REVUE R.A.I.R.O. nº Février 1973 - J-1, Pages 3 à 54

NEW METHODS FOR NONLINEAR FILTERING (1)

R.S. BUCY (2), C. HECHT (3), K.D. SENNE (4)

Résumé. — *Cet article concerne la réalisation de filtres non linéaires, utilisant un calculateur numérique de la troisième génération comme moyen de synthèse.*

On considère en particulier la synthèse et l'évaluation par une méthode de Monte-Carlo d'un démodulateur de phase optimal. Les performances du démodulateur optimal sont comparées avec celles de la « boucle à verrouillage de phase » classique et on montre qu'une amélioration de performance sur l'erreur de 2 à 3 db est obtenue. Finalement on indique des méthodes pour la synthèse pratique du filtre optimal qui font actuellement l'objet de recherches.

I - INTRODUCTION

One of the basic problems of experimental science consists of processing noisy observations (non linear functions of a subset of some underlying set of state variables, corrupted by measurement noise) to obtain estimates of *all* of the state variables. Usually the state variables are adequately modelled as solutions to a set of nonlinear stochastic differential equations, and the observation mechanism can presumably be modeled by an associated system of stochastic differential equations whereby the measurements are taken to be nonlinear functions of the states with additive noises. The nonlinear problem is of course a natural generalization of the linear filtering problem which has been completely resolved by the Kalman-Bucy filtering theory [7], [12]. Both filtering problems are concerned with finding a « best » estimate (i.e.,

(1) This research was supported in part by the United States Air force (Air Force Systems Command) under Air Force Office of Scientific Research grant AFOSR - 71-2141 and under Frank J. Seiler Research Laboratory project 7904, as well as in part by NASA under grant NAS5-10789.

(2) University of Southern California, Los Angeles U.S.A. and Université Paul Sabatier and Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS Toulouse (France).

(3) T.R.W. Systems, Redondo Beach, California, U.S.A.

(4) M.I.T. Lincoln Laboratory, Lexington, Massachusetts, U.S.A. Formerly with the Frank J. Seiler Research Laboratory, Air Force systems Command U.S. Air Force Academy, Colorado U.S.A.

an estimate minimizes the expected value of some loss function of the estimation error), and, consequently, are resolved by computing the conditional density of the present states given the past and present observations of the measurement process. The theorical description of the evolution of the conditional densities is contained in the " Representation theorem " [12]; however a challenging numerical problem remains : construct *a finite-dimensional algorithm which accepts as inputs the observations and produces as outputs the optimal estimates.* The numerical problem, referred to as the " realization problem ", is the subject of the present paper.

This paper is a natural sequel to [7], in which both the linear and nonlinear filtering problems were reviewed and the results put in perspective. Some of the early work on the realization problem [15], [31] was mentioned at that time. In the three intervening years since the survey in [7] appeared, much work has been reported involving the realization problem [2], [3], [8], [16], [22], [24], [35], [38], [41] and [43]. An up to date survey of the more recent research, including the present results has been included in the F.J. Seiler Laboratory report [10] and part of the results of that report are reproduced in this paper in order that they may be exposed to a larger audience in a more formal way.

Our research is described in this paper in relation to a specific application to a problem concerning the demodulation of a phase-modulated carrier corrupted by additive noise, where the phase process is taken as doubly-integrated white noise. It is necessary to confine attention to a single class of applications for realization studies since unlike the linear counterpart there is no uniformly, best finite-dimensional realization for nonlinear filters. In general each application has features which must be exploited, for example, the natural domain for the phase demodulator states is a torus.

Even for a specific application we are obliged to select among a variety of approximation techniques as a function of problem parameters; for example, the signal to noise ratio is a critical variable in the current application. Naturally, we are concerned with developing generally applicable techniques, and we therefore emphasize the rationale for selecting the various approximation methods.

The phase demodulation problem possesses three intriguing features : first it is technological extremely important (for a survey of the astounding number of attempts to extend the threshold of the classical phase-locked loop see the references in [31]); secondly, the state dimension of the signal process is only two, thereby making feasible Monte Carlo simulations on current serial third-generation digital computers; and, finally the two-state model for the signal process is realistic. Generalization of methods to problems of higher signal state dimension is possible but costly in terms of computation time on current digital computers. In the final section of this paper we discuss some synthesis tools which we feel will be a more effective substitute for serial digital computers.

The research in this paper is an attempt to use the theory of nonlinear filtering to design a real system. In the authors' opinion too often research in nonlinear filtering, can be classified into two categories : either it is solely concerned with theory or it consists of engineering fixes to the extended Kalman-Bucy filters or moment filters. While both of these categories are useful, we believe our approach offers considerable fresh possibilities both for developing the theory and for designing practical systems. For without practical examples of nonlinear filters, the theory is likely to diverge from significant and useful questions, while engineering design techniques tend to degenerate to stale and unimaginative cut and try cookbook methods; in a word, practical examples develop insight.

II - OPTIMAL NONLINEAR PHASE DEMODULATION

1. Introduction

An interesting application for optimal nonlinear estimation was introduced by Mallinckrodt, Bucy, and Chang [31], who considered the problem of tracking a first-order phase process based on measurements of a modulated signal in noise of the form:

$$dz(t) = A \cos [\omega_0 t + x_1(t)]\, dt + dv(t)$$

where A is a known amplitude, ω_0 is a known carrier frequency, and $x_1(t)$ is the message process being tracked. The measurement noise is assumed white. Using a voltage-controlled oscillator the known carrier may be removed by heterodyning down to base band, producing both in-line and quadrature components and resulting in an equivalent two-dimensional measurement process of the form :

$$\begin{bmatrix} dz_1(t) \\ dz_2(t) \end{bmatrix} = \begin{bmatrix} \cos x_1(t) \\ \sin x_1(t) \end{bmatrix} dt + \begin{bmatrix} dv_1(t) \\ dv_2(t) \end{bmatrix} \tag{1}$$

where A has been taken as unity without loss of generality, and the noise has been replaced by a vector of mutually independent quantities.

The first-order phase process studied in [31] consisted of Brownian motion with increment of length h having variance qh. In this paper we describe a study of a second order phase process involving the integral of Brownian motion, expressed as :

$$\begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} d\beta_t \tag{2}$$

We will retain the same measurement model (1) and let the noises v_1 and v_2 be independent Brownian motions with path increments of length h having variance rh.

The familiar technique for tracking such phase processes involves the application of the phase-locked loop, studied thoroughly by Viterbi [8]. The phase-locked loop is a very ingenious nonlinear estimator, capable of near-optimal performance in good signal/noise environments. The steady-state behavior of the loop is identical to that of the so-called " linearized " or " extended " Kalman-Bucy filter for nonlinear systems, however, as was illustrated in [31]. Thus, the phase-locked loop is an excellent example of a very successful extended Kalman-Bucy filter. In the following section we discuss the stationary behavior of the linearized filter and the consequences of time descretization of the problem. Then in the subsequent section we will recast the discrete-time solution in terms of optimal nonlinear estimation, thereby setting the stage for a description of the numerical experiments.

2. The Linearized Kalman-Bucy Filter

Equations for the standard, continuous, linearized Kalman-Bucy filter are reproduced in Table 1 for the above phase-estimation problem. The measurement function is linearized about the current estimate $\hat{x}_1(t)$ of the phase. The approximate filter attempts to track the mean of the conditional phase density, which, of course, is the minimum mean-squared error estimate. As a result, the loop estimate takes on all real values, whereas the original problem is only observable modulo 2 π. Accordingly, it makes sense to consider the modulus of the error in the interval $[-\pi, \pi]$, or equivalently to take $E[(e + \pi) \bmod 2\pi - \pi]^2$ as an error criterion. Naturally if the signal to noise ratio is high enough we would expect the minimum of the mean-modulo-2 π- squared to be essentially equivalent to mean-squared error, but for higher noise situations, the modulation of the error would tend to bound the maximum mean-modulo-2 π-squared error (since the worst case will be a uniform error density on $[-\pi, \pi]$). In Section C below we will discuss a nonlinear estimate designed to be defined only on $[-\pi, \pi]$.

Table 1
Summary of Continuous Linearized Kalman-Bucy Filter

Phase Process Model

$$d\underline{x} = F\underline{x}dt + G\,d\beta, \quad \underline{x}(o) \sim N[\underline{o}, \textstyle\sum(o)]$$

$$F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad E(\beta_t - \beta_0)^2 = \int_0^t q\,ds.$$

Observation Model

$$\mathrm{d}\underline{z} = h(\underline{x})\,\mathrm{d}t + \mathrm{d}\underline{v}, \quad E(\underline{v}_t - \underline{v}_0)(\underline{v}_t - \underline{v}_0)' = \int_0^t R\,\mathrm{d}s,$$

$$R = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}, \quad h(\underline{x}) = \begin{bmatrix} \cos x_1 \\ \sin x_1 \end{bmatrix}.$$

Filter Model

$$\mathrm{d}\hat{\underline{x}} = (F - KH)\,\hat{\underline{x}}\mathrm{d}t + K(\mathrm{d}\underline{z} - \mathrm{d}\hat{\underline{z}} + H\hat{x}) \tag{4}$$

$$= F\hat{\underline{x}}\mathrm{d}t + K(\mathrm{d}z - \mathrm{d}\hat{z}).$$

$$\mathrm{d}\hat{\underline{z}} = h(\hat{\underline{x}})\,\mathrm{d}t, \quad H = \nabla_x h(\hat{\underline{x}}) = \begin{bmatrix} -\sin\hat{x}_1 & 0 \\ \cos\hat{x}_1 & 0 \end{bmatrix},$$

$$K = PH'R^{-1} = \begin{bmatrix} -\dfrac{P_{11}}{r}\sin\hat{x}_1 & \dfrac{P_{11}}{r}\cos\hat{x}_1 \\ -\dfrac{P_{12}}{r}\sin\hat{x}_1 & \dfrac{P_{12}}{r}\cos\hat{x}_1 \end{bmatrix},$$

$$\dot{P} = FP + PF' - PH'R^{-1}HP + Q \tag{5}$$

$$= \begin{bmatrix} \dot{P}_{11} & \dot{P}_{12} \\ \dot{P}_{12} & \dot{P}_{22} \end{bmatrix} = \begin{bmatrix} 2P_{12} - \dfrac{P_{11}^2}{r} & P_{22} - \dfrac{P_{11}P_{12}}{r} \\ P_{22} - \dfrac{P_{11}P_{12}}{r} & q - \dfrac{P_{12}^2}{r} \end{bmatrix}$$

Equilibrium Solution

$$P = \begin{bmatrix} \sqrt{2}\,r^{3/4}q^{1/4} & r^{1/2}q^{1/2} \\ r^{1/2}q^{1/2} & \sqrt{2}\,r^{1/4}q^{3/4} \end{bmatrix} = r^{1/2}q^{1/2}\begin{bmatrix} \tau & 1 \\ 1 & \dfrac{2}{\tau} \end{bmatrix} \tag{6}$$

Filter Time Constant

$$\tau = \sqrt{2}\left(\frac{r}{q}\right)^{1/4}$$

In order to implement the nonlinear filter on a digital computer we will need to discretize time by some interval Δ. The selection of Δ will be made so as to assure essentially the same steady-state performance of the continuous and discrete phase-locked loops.

Accordingly, we will now evaluate the steady-state solution of the matrix Ricatti-equation for the linearized filter, which we know is approximately equal to the error variance for low-noise applications. If we make the definitions :

$$F \triangleq \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad Q \triangleq \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix},$$

$$H \triangleq \begin{bmatrix} -\sin \hat{x}_1 & 0 \\ \cos \hat{x}_1 & 0 \end{bmatrix}, \quad R \triangleq \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}, \tag{3}$$

then the linearized filter satisfies the equations

$$\mathrm{d}\underline{\hat{x}} = (F - KH)\,\underline{\hat{x}}\mathrm{d}t + K(\mathrm{d}\underline{z} - \mathrm{d}\underline{\hat{z}} + H\underline{\hat{x}}\mathrm{d}t) \tag{4}$$

where

$$\mathrm{d}\underline{\hat{z}} = \begin{bmatrix} \cos \hat{x}_1 \\ \sin \hat{x}_1 \end{bmatrix} \mathrm{d}t, \quad \underline{\hat{x}}(0) = 0$$

and

$$K = PH'\,R^{-1},$$

where P satisfies

$$\frac{\mathrm{d}P}{\mathrm{d}t} = FP + PF' - PH'\,R^{-1}\,HP + Q, \tag{5}$$

with $P(0) = \sum$. This system is diagrammed in Figure 1. We may set the derivative in (5) to zero as a necessary condition for steady-state and algebraically determine the possible steady-state solutions. Accordingly, we obtain

$$P = \begin{bmatrix} \sqrt{2}\, r^{3/4} q^{1/4} & r^{1/2} q^{1/2} \\ r^{1/2} q^{1/2} & \sqrt{2}\, r^{1/4} q^{3/4} \end{bmatrix}$$

$$= \sqrt{rq} \begin{bmatrix} \sqrt{2} \left(\frac{r}{q}\right)^{1/4} & 1 \\ 1 & \sqrt{2} \left(\frac{q}{r}\right)^{1/4} \end{bmatrix} \tag{6}$$

$$= \sqrt{rq} \begin{bmatrix} \tau & 1 \\ 1 & \frac{2}{\tau} \end{bmatrix}, \tag{7}$$

where $\tau = \sqrt{2} \left(\frac{r}{q}\right)^{1/4}$, the filter time constant, is obtained by studying the $\mathrm{d}\underline{\hat{x}}$

equation (4) with (6) substituted for P as follows : The homogeneous part of (4) is rewritten as :

$$\begin{aligned} d\underline{\hat{x}} &= (F - KH)\,\underline{\hat{x}}\,dt \\ &= (F - PH'\,R^{-1}\,H)\,\underline{\hat{x}} \\ &= \begin{bmatrix} -\frac{1}{r}\sqrt{2\,r^{3/2}\,q^{1/2}} & 1 \\ -\frac{1}{r}\sqrt{rq} & 0 \end{bmatrix} \underline{\hat{x}}\,dt. \end{aligned} \tag{8}$$

The eigenvalues of the matrix in (8) are the solutions λ to

$$\lambda^2 + \frac{1}{r}\sqrt{2\,r^{3/2}\,q^{1/2}}\,\lambda + \frac{1}{r}\sqrt{rq} = 0, \tag{9}$$

resulting in $\lambda = \frac{1}{\sqrt{2}}\left(\frac{q}{r}\right)^{1/4}(-1 \pm i)$, or, using the definition for τ in (7),

$$\lambda = \frac{1}{\tau}(-1 \pm i). \tag{10}$$

Thus the solutions to (8) are of the form :

$$\hat{x}_t = C_0\,e^{-\frac{t}{\tau}}\cos\left(\frac{t}{\tau} + C_1\right),$$

so that τ is indeed a time constant. From (4) and (8) we may write the steady-state differential equations for the extended Kalman-Bucy filter (i.e., the phase-locked loop) as :

$$d\underline{\hat{x}} = F\hat{x}\,dt + K(d\underline{z} - d\underline{\hat{z}}), \tag{11}$$

or, equivalently,

$$d\hat{x}_1 = \hat{x}_2\,dt + \frac{2}{\tau}(-\sin\hat{x}_1\,dz_1 + \cos\hat{x}_1\,dz_2),$$

and

$$d\hat{x}_2 = \frac{2}{\tau^2}(-\sin\hat{x}_1\,dz_1 + \cos\hat{x}_1\,dz_2). \tag{12}$$

Refer to Table 1 for a summary of above results. The corresponding results for the discrete-time filter, obtained by Hecht [24], are summarized in Table 2. If we are interested in simulating a filter which behaves substantially the same as the continuous filter then we must choose the sampling interval carefully so that it is as large as possible subject to the constraint that the discrete covariance in steady-state is within an acceptable tolerance of the continuous covariance.

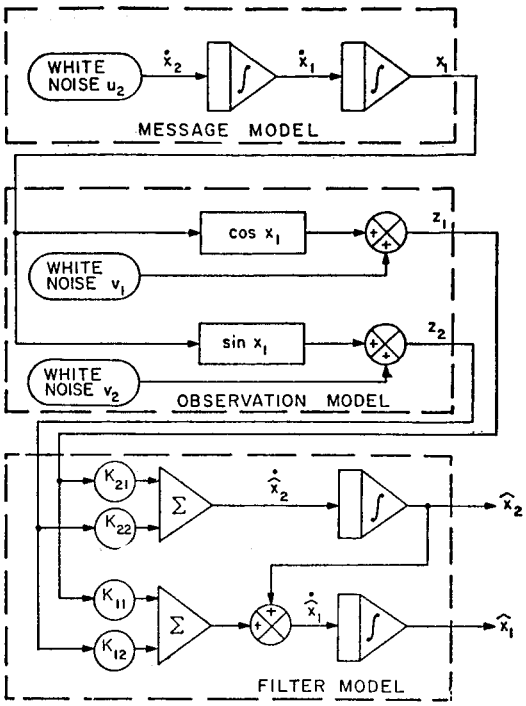


Figure 1
Block Diagram of Linearized Phase Estimation

Table 2
Summary of Discrete Linearized Kalman-Bucy Filter

Phase Process Model

$$\underline{x}(n\Delta) = \Phi\,[n\Delta, (n-1)\Delta]\,\underline{x}\,[(n-1)\Delta] + \Gamma \underline{u}_d(n\Delta),$$

$$\Phi\,[n\Delta, (n-1)\Delta] = e^{F\Delta} \cong \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix} \triangleq \Phi$$

$$E[u_d^2(n\Delta)] = q_d = \frac{q}{\Delta}, \quad n\Delta = t$$

$$\Gamma = \int_{(n-1)\Delta}^{n\Delta} \Phi\,[n\Delta, \mu]\, G\, d\mu \cong G\Delta = \begin{bmatrix} 0 \\ \Delta \end{bmatrix}$$

Observation Model

$$\underline{z}(n\Delta) = \underline{h}\,[\underline{x}(n\Delta)] + \underline{v}_d(n\Delta)$$

$$h\,[\underline{x}(n\Delta)] = \begin{bmatrix} \cos x_1(n\Delta) \\ \sin x_1(n\Delta) \end{bmatrix}$$

$$E\,[\underline{v}_d(n\Delta)\underline{v}'_d(n\Delta)] = R_d = \begin{bmatrix} \frac{r}{\Delta} & 0 \\ 0 & \frac{r}{\Delta} \end{bmatrix}$$

Predictor Model

$$\underline{\hat{x}}(n \mid n-1) = \Phi\underline{\hat{x}}(n \mid n)\,,\ \underline{\hat{x}}(o \mid o) = \underline{o}$$

$$S(n) = \Phi P_d(n-1)\Phi' + \Gamma Q_d \Gamma'$$

$$= \begin{bmatrix} S_{11}(n) & S_{12}(n) \\ S_{12}(n) & S_{22}(n) \end{bmatrix}$$

Filter Model

$$\underline{\hat{x}}(n \mid n) = \underline{\hat{x}}(n \mid n-1) + A(n)\,[\underline{z}(n) - \underline{\hat{z}}(n \mid n-1)]$$

$$\underline{\hat{z}}(n \mid n-1) = h\,[\underline{\hat{x}}(n \mid n-1)],$$

$$A(n) = S(n)H'\,[HS(n)H' + R_d]^{-1}$$

$$= \frac{1}{S_{11} + r_d} \begin{bmatrix} -S_{11}(n)\sin\hat{x}_1(n \mid n-1) & S_{11}(n)\cos\hat{x}_1(n \mid n-1) \\ -S_{12}(n)\sin\hat{x}_1(n \mid n-1) & S_{12}(n)\cos\hat{x}_1(n \mid n-1) \end{bmatrix}$$

$$P_d(n) = [I - A(n)H]\,S(n)\,[I - A(n)\,H]' + A(n)R_dA'(n)$$

$$= \frac{1}{S_{11} + r_d} \begin{bmatrix} S_{11}(n) & S_{12}(n) \\ S_{12}(n) & S_{22}(n)\left(\frac{S_{11}(n) + r_d}{r_d}\right) - \frac{S_{12}^2(n)}{r_d} \end{bmatrix}$$

Equilibrium Solution

$$S_{11}(n+1) = \frac{r_d\,[S_{11}(n) + 2\,S_{12}(n)\Delta] - S_{12}^2(n)\Delta^2}{S_{11}(n) + r_d} + S_{22}(n)^2$$

$$S_{12}(n+1) = \frac{S_{12}(n)\,[r_d - S_{12}(n)\Delta]}{S_{11}(n) + r_d} + S_{22}(n)\Delta$$

$$S_{22}(n+1) = \frac{-S_{12}^2(n)}{S_{11}(n) + r_d} + S_{22}(n) + \Delta^2 q_d$$

Setting $S(n+1) = S(n) = S$, using, a discrete form of the Bass-Roth Theorem [1],

$$S = \begin{bmatrix} r_d\left(\frac{1}{\alpha_0} - 1\right) & \frac{\Delta\sqrt{q_d r_d}}{\sqrt{\alpha_0}} \\ \frac{\Delta\sqrt{q_d r_d}}{\sqrt{\alpha_0}} & \sqrt{q_d r_d}\left(\frac{1}{\sqrt{\alpha_0}} - \sqrt{\alpha_0}\right) + \Delta^2 q_d \end{bmatrix}$$

$$= \begin{bmatrix} \frac{r}{\Delta}\left(\frac{1}{\alpha_0} - 1\right) & \frac{\sqrt{qr}}{\sqrt{\alpha_0}} \\ \frac{\sqrt{qr}}{\sqrt{\alpha_0}} & \frac{\sqrt{qr}}{\Delta}\left(\frac{1}{\sqrt{\alpha_0}} - \sqrt{\alpha_0}\right) + \Delta q \end{bmatrix},$$

with

$$\alpha_0 = 1 + \frac{\rho}{4} + \frac{m^{1/2}}{2} - \frac{1}{2}\left[\frac{\rho^2}{2} + 6\rho + (4+\rho)m^{1/2}\right]^{1/2}$$

$$m = \frac{\rho^2}{2} + 4\rho$$

$$\rho = \frac{q}{r}\Delta^4$$

The steady-state discrete prediction covariance S and the filtering covariance P_d must converge to the continuous P of (6) in the limit as $\Delta \to 0$. Accordingly, we write S as a function of P and Δ/τ as :

$$S_{11}\left(\frac{\Delta}{\tau}\right) = P_{11}\frac{\tau}{2\Delta}\left(\frac{1}{\alpha_0} - 1\right), \tag{13}$$

$$S_{12}\left(\frac{\Delta}{\tau}\right) = P_{12}\frac{1}{\sqrt{\alpha_0}}, \tag{14}$$

$$S_{22}\left(\frac{\Delta}{\tau}\right) = P_{22}\left[\frac{\tau}{2\Delta}\left(\frac{1}{\sqrt{\alpha_0}} - \sqrt{\alpha_0}\right) + \frac{\Delta}{\tau}\right], \tag{15}$$

where α_0 is a function of $\frac{\Delta}{\tau}$, since the ρ of Table 2 may be written :

$$\rho\left(\frac{\Delta}{\tau}\right) = 4\left(\frac{\Delta}{\tau}\right)^4. \tag{16}$$

The filtering steady state P_d may be similarly expressed as :

$$P_{d_{11}}\left(\frac{\Delta}{\tau}\right) = \alpha_0 S_{11}\left(\frac{\Delta}{\tau}\right) = P_{11}\frac{\tau}{2\Delta}(1 - \alpha_0), \tag{17}$$

$$P_{d_{12}}\left(\frac{\Delta}{\tau}\right) = \alpha_0 S_{12}\left(\frac{\Delta}{\tau}\right) = P_{12}\sqrt{\alpha_0}, \tag{18}$$

$$P_{d_{22}}\left(\frac{\Delta}{\tau}\right) = S_{22}\left(\frac{\Delta}{\tau}\right) - \frac{S_{12}^2\left(\frac{\Delta}{\tau}\right)}{S_{11}\left(\frac{\Delta}{\tau}\right) + r_d}$$

$$= P_{22}\left[\frac{\tau}{2\Delta}\left(\frac{1}{\sqrt{\alpha_0}} - \sqrt{\alpha_0}\right) + \frac{\Delta}{\tau}\right] - \frac{P_{12}^2}{P_{11}}(1 - \alpha_0) \tag{19}$$

The relationship between Δ/τ and the discrete variances is illustrated in figures 2-4. If we choose $\Delta/\tau = 0.1$ and evaluate (13) and (17), we find :

$$\frac{S_{11}(0.1)}{P_{11}} = 1.108028,$$

while

$$\frac{P_{d_{11}}(0.1)}{P_{11}} = 0.9070263,$$

or $S_{11}(0.1) - P_{d_{11}}(0.1) \cong 0.2\, P_{11}$. Thus we suffer a 10 % change in the steady-state covariance by taking 10 samples per time constant. In any case we will compare all discrete filters to each other, and we can assume that results lie within 10 % of their continuous limits.

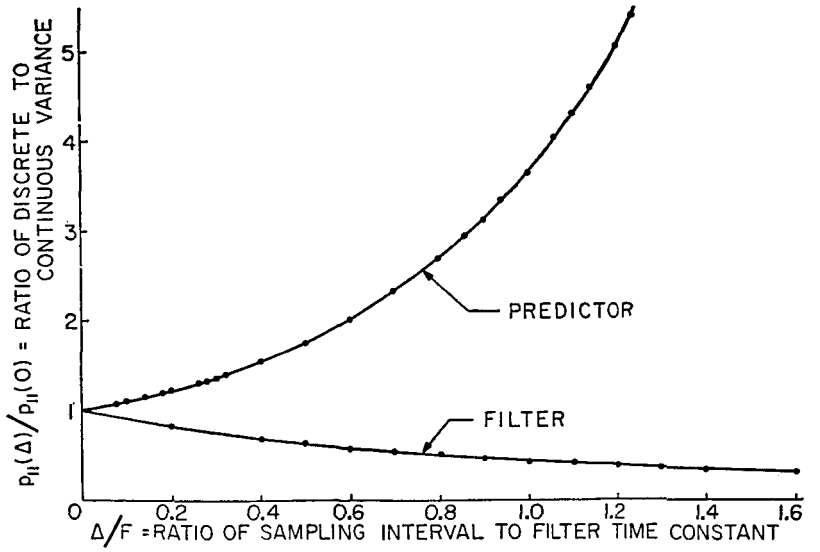


Figure 2
Discrete P_{11} Error Variance

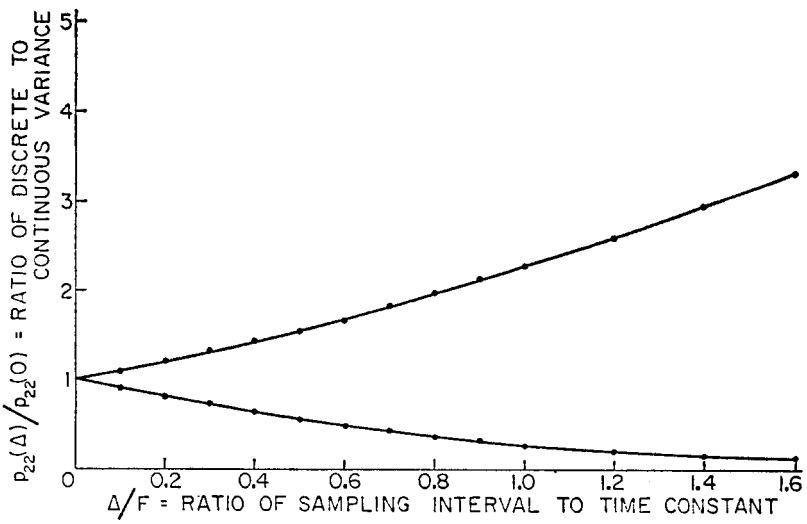


Figure 3
Discrete P_{22} Error Variance

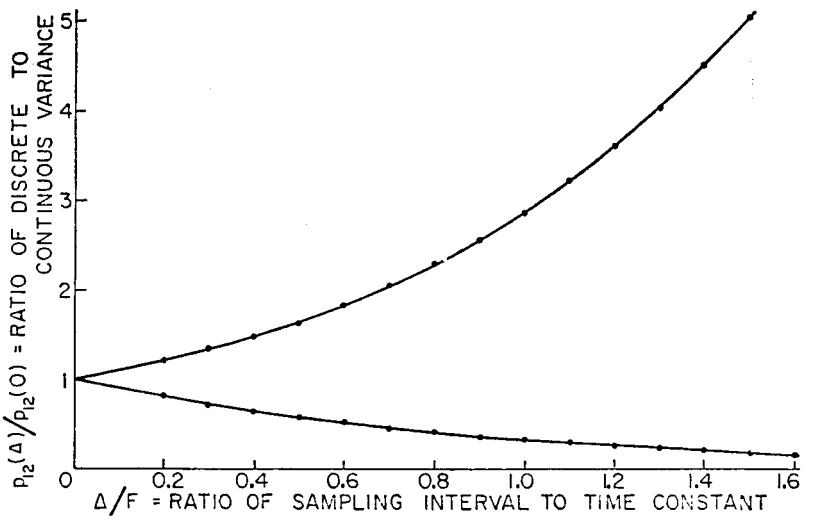


Figure 4
Discrete P_{12} Error Variance

3. Application of Nonlinear Filtering

The application of Bayes-Law filtering, developed by Bucy and Senne [14], requires special considerations for each special case. In the problem of this paper, for example, the dimension of the driving noise is one less than that of the state vector, a situation for which Bucy and Senne indicate there results in a computational simplification for Bayes Law implementation. We will explore that claim here for the second order phase process.

At the outset consider a similar but unrealistic phase process with two driving noises u_1 and u_2 so that (in discrete time) :

$$\underline{x}(n+1) = \Phi\underline{x}(n) + \Gamma\underline{u}(n), \tag{20}$$

where

$$\Gamma = \begin{bmatrix} \varepsilon\Delta & 0 \\ 0 & \Delta \end{bmatrix}, \quad \underline{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

and the remaining quantities, as well as the sensor model remain the same as described in Table 2. Now if we let the variance of u_2 and u_1 be q/Δ and take the limit as $\varepsilon \to 0$ then the process (20) will be identical to the one described in Table 2. For any finite ε, however, the probability density function of the driving terms $\Gamma\underline{u}$ will take the form :

$$P_{\Gamma\underline{u}}(\underline{\xi}) = \frac{1}{2\pi \det [A]} \exp\left\{ -\frac{1}{2} \| \underline{\xi} \|^2_{A^{-1}} \right\}, \tag{21}$$

where

$$A \triangleq \Gamma Q \Gamma' = \begin{bmatrix} \varepsilon^2\Delta q & 0 \\ 0 & \Delta q \end{bmatrix}.$$

Similarly, the density of the observation noise will be denoted $P_v(\underline{\xi})$, which will be gaussian with covariance R.

If we review the Bayes representation theorem solution to the discrete filtering problem [12], we can determine that :

$$J_{n+1|n}(\underline{y}) = \frac{1}{K_n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{\Gamma\underline{u}}(\underline{y} - \Phi\underline{x}) J_{n|n}(\underline{x}) dx_1 dx_2, \tag{22}$$

and

$$J_{n|n}(\underline{y}) = \frac{1}{K_n} P_v [\underline{z}_n - \underline{h}(\underline{y})] \, J_{n|n-1}(\underline{y}) \tag{23}$$

Now if we take the limit as $\varepsilon \to 0$ in (22) we obtain

$$J_{n+1|n}(\underline{y}) = \frac{C}{K_n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta\,[y_1 - x_2\Delta - x_1] \exp\left\{ -\frac{1}{2\,q\Delta}\,(y_2 - x_2)^2 \right\} J_{n|n}(\underline{x}) \mathrm{d}x_1 \mathrm{d}x_2$$

$$= \frac{C}{K_n} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\,q\Delta}\,(y_2 - x_2)^2 \right\} J_{n|n} \begin{pmatrix} y_1 - x_2\Delta \\ x_2 \end{pmatrix} \mathrm{d}x_2, \qquad (24)$$

where $C = (2\,\pi)^{-1/2}(q\Delta)^{-1/2}$. In other words the Bayes integral reduces to an integral over a subspace of the state space with dimension equal to that of the driving noise vector (equals one in this case). The complete Bayes recursion may now be written :

$$J_{n+1|n}(\underline{y}) = C_1 \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\,q\Delta}\,(y_2 - x_2)^2 \right\} J_{n|n} \begin{pmatrix} y_1 - x_2\Delta \\ x_2 \end{pmatrix} \mathrm{d}x_2, \quad (25)$$

$$J_{n|n}(\underline{y}) = C_2 \exp\left\{ -\frac{\Delta}{2\,r}\,[(z_1 - \cos y_1)^2 + (z_2 - \sin y_1)^2] \right\} J_{n|n-1}(\underline{y}) \quad (26)$$

where both C_1 and C_2 are normalizing constants.

Although there can be no doubt that a significant simplification has resulted from reducing the number of integrations required for the Bayes integral computation, it must be pointed out that the arguments of $J_{n|n}$ in (25) must be determined to lie in a particular subspace of R^2. If the densities have been stored only at a finite number of points then some form of interpolation of the densities will be necessary, resulting in some overhead, so that the computation is not exactly equivalent to a scalar problem.

Another problem of concern to this particular application is the domain of the conditional probability densities. If we are only interested in modulo-2 π errors for the reasons stated earlier then there is no loss of generality to map all modulo-2 π intervals of the conditional densities back into the $[-\pi, \pi]$ interval in the x_1 coordinate and sum up the individual contributions. Similarly, because we are computing the phase in discrete-time with sample time Δ, we observe that phase-rate contributions outside the interval $\left[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}\right]$ will have the same effect on the next phase as their modulo $-\frac{2\,\pi}{\Delta}$ component.

Accordingly, we may map the phase-rate x_2 components of the conditional densities back into $\left[\frac{-\pi}{\Delta}, \frac{\pi}{\Delta}\right]$ for the same reason as above. The combination of the two mappings results in an equivalence between the " cyclic " state space and a torus (see fig. 5).

Next we consider what simplifications arise for the Bayes integral update (25)-(26) as a consequence of the cyclic mapping of the state space. We begin by combining (25) and (26) into a single equation representing the filter update, and absorb all non state-dependent terms of the quadratic exponent expansions expansions into a single normalizing constant C_0, giving :

$$J_{n+1|n+1}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = S_{n+1}(y_1) \int_{-\infty}^{\infty} \exp\left\{ -\frac{(y_2 - \mu)^2}{2\,q\Delta} \right\} J_{n|n}\begin{pmatrix} y_1 - \mu\Delta \\ \mu \end{pmatrix} d\mu, \quad (27)$$

where

$$S_{n+1}(y_1) \triangleq C_0 \exp\left\{ \frac{z_1(n+1)\cos y_1 + z_2(n+1)\sin y_1}{r/\Delta} \right\}. \quad (28)$$

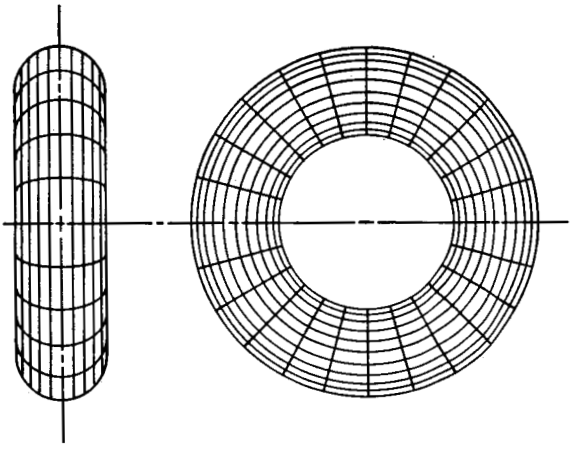


Figure 5
Torus Interpretation of Doubly Cyclic State Space

If we now modulate the density $J_{n|n}$ as described above, we obtain :

$$\tilde{J}_{n|n}\begin{pmatrix} \sigma \\ \tau \end{pmatrix} \triangleq \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} J_{n|n}\begin{pmatrix} \sigma + 2\pi k \\ \tau + \dfrac{2\pi\, 1}{\Delta} \end{pmatrix}, \quad (29)$$

with $-\pi \leqslant \sigma < \pi$, and $\dfrac{-\pi}{\Delta} \leqslant \tau < \dfrac{\pi}{\Delta}$. Finally we use the definition (29) on $J_{n+1|n+1}$, and substitute (27), resulting in the following manipulations [9] :

$$\tilde{J}_{n+1|n+1}\begin{pmatrix} \sigma \\ \tau \end{pmatrix} = \sum_k \sum_l S_{n+1}(\sigma + 2\pi k) \int_{-\infty}^{\infty} \exp\left\{ \frac{-(\tau + 2\pi\, 1/\Delta - \mu)^2}{2\,q\Delta} \right\} J_{n|n}\begin{pmatrix} \sigma - \mu\Delta + 2\pi k \\ \mu \end{pmatrix} d\mu \quad (30)$$

$$= S_{n+1}(\sigma) \sum_k \sum_l \sum_m \int_{\frac{-\pi}{\Delta}+\frac{2\pi m}{\Delta}}^{\frac{\pi}{\Delta}+\frac{2\pi m}{\Delta}} \exp\left\{\frac{-(\tau + 2\pi 1/\Delta - \mu)^2}{2q\Delta}\right\} J_{n|n}\begin{pmatrix}\sigma - \mu\Delta + 2\pi k \\ \mu\end{pmatrix} d\mu$$

$$\left(\text{let } \xi = \mu - \frac{2\pi m}{\Delta}\right)$$

$$= S_{n+1}(\sigma) \sum_k \sum_l \sum_m \int_{\frac{-\pi}{\Delta}}^{\frac{\pi}{\Delta}} \exp\left\{\frac{-[\tau - \xi + 2\pi(1-m)/\Delta]^2}{2q\Delta}\right\} J_{n|n}\begin{pmatrix}\sigma - \xi\Delta + 2\pi(k-m) \\ \xi + \frac{2\pi m}{\Delta}\end{pmatrix} d\xi$$

(let $i = l - m, j = l - i$)

$$= S_{n+1}(\sigma) \sum_i \sum_j \sum_k \int_{\frac{-\pi}{\Delta}}^{\frac{\pi}{\Delta}} \exp\left\{\frac{-(\tau - \xi + 2\pi i/\Delta)^2}{2q\Delta}\right\} J_{n|n}\begin{pmatrix}\sigma - \xi\Delta + 2\pi(k-j) \\ \xi + \frac{2\pi j}{\Delta}\end{pmatrix} d\xi$$

(Fubini's theorem to interchange $\sum$ and $\int$)

$$= S_{n+1}(\sigma) \int_{\frac{-\pi}{\Delta}}^{\frac{\pi}{\Delta}} a(\tau - \xi)\, \tilde{J}_{n|n}\begin{pmatrix}\sigma - \xi\Delta \\ \xi\end{pmatrix} d\xi, \tag{31}$$

where

$$a(\tau - \xi) \triangleq \sum_{i=-\infty}^{\infty} \exp\left\{\frac{-(\tau - \xi + 2\pi i/\Delta)^2}{2q\Delta}\right\} \tag{32}$$

The result (31) represents the exact recurrence relation required for the cyclic conditional density. The individual terms of $a(\cdot)$ will drop rapidly on either side of $\tau - \xi$, depending only on the variance $q\Delta$.

Having constructed a density function updating formula for the phase estimation problem, the question remains — what form shall the phase estimate itself take ? It is clear that the conditional mean (the goal of the phase-locked loop) is an admissible candidate. The cost criterion that is minimized by the conditional mean, however, is the mean squared error. It is not obvious that mean squared error is the best criterion for choosing estimates modulo 2π. In fact, a periodic cost function of the form :

$$L(e) = 2(1 - \cos e) \tag{33}$$

might be more appropriate than e^2 if only modulo-2 π-errors are important. The cyclic loss (33) looks like e^2 for small e and $(e - 2k\pi)^2$ for e close to $2k\pi$ for all k. Moreover, we may easily show that the estimate x_n^* which minimizes (1) :

$$E\,[L(x_n^* - x_n^1) \mid Z_n] = \int_{-\pi}^{\pi} \int_{\frac{-\pi}{\Delta}}^{\frac{\pi}{\Delta}} L(\tau - x_n^*)\, \tilde{J}_n(\tau, \sigma)\, d\sigma\, d\tau \qquad (34)$$

is given by

$$x_n^* = \tan^{-1} \{ E\,[\sin(x_n) \mid Z_n]/E\,[\cos(x_n) \mid Z_n] \} \qquad (35)$$

For a proof of (35) see [31].

Of course the cyclic loss may not be the only desirable loss function for the phase demodulator. But many other proposed criteria would be minimized by appropriate function of the conditional phase densities, resulting in extremely great flexibility for the experiment designer. Moreover, it is seen that the conditional expected loss (34) as well as the densities themselves are computable in addition to the estimates so that a considerable amount of quantitative information is available to provide a realistic assessment of the quality of the estimates, regardless of the cost criterion employed. No such information is provided by the phase-locked loop — especially after steady-state is attained.

Appendix A. Numerical Experiments With The Phase-Locked Loop

In order to provide an accurate check on the value of the nonlinear filters, it was necessary to perform extensive Monte-Carlo tests on the phase-locked loop (i.e. the steady-state linearized filter). Since the discrete phase-locked loop operates very fast (over 1,000 estimates per second on the CDC 6600), it was possible to average estimates over 5000 sample paths of length 130 in 10 mn. If three time constants are discarded (30 samples) in each sample path the resulting 100 estimates represent steady-state. If all of the steady-state errors were averaged this would lead to 500,000 monte-carlos of the steady-state error. On the other hand, since adjacent errors are correlated, the effective Monte-Carlo length would better be set between $N = 50{,}000$ (one estimate per time constant) and $N = 500{,}000$ (every estimate included). Thus, we may determine the three-standard deviation confidence bands based on both values of N.

The independent parameters for the different phase-locked loop cases considered was p_{11}, the steady-state continuous Ricatti equation solution for the phase error variance. $p_{11} = \sqrt{2}\, r^{3/4} q^{1/4}$ is shown by Viterbi [42] to be the inverse of the effective signal to noise ratio, or N/S. The initial condition for the matrix

(1) Z_n denotes the data up to and including time n,

$$z_1(n),\ z_2(n),\ z_1(n-1),\ z_2(n-1),\ \ldots,\ z_1(0),\ z_2(0)$$

Ricatti equation can be taken to be its steady-state value (1), and the mean-squared error will then only be a function of N/S and not of q. A value of 0.01 was chosen for q, and thus the value of r was taken to be $(p_{11})^{4/3}/(2^{2/3} q^{1/3})$.

Now if the phase-locked loop were linear with $H = [-1 \quad 0]$, then p_{11} would be the steady-state mean-squared error. An equivalent interpretation would be to consider p_{11} as the steady-state mean-squared error of a filter operating on the linear measurements $z = Hx + y$. The latter interpretation

Table A-1
Confidence Intervals for Linearized Filter

N/S (dB)	μ_2	μ_4	$\rho_2 = \frac{\mu_4}{\mu_2}$	Confidence (dB)	
				$N = 50{,}000$	$N = 500{,}000$
— 14.0	.0372	.0372	3.06	± 0.0828	± 0.0264
— 12.0	.0599	.0114	3.16	± 0.0848	± 0.0270
— 10.0	.0993	.0321	3.26	± 0.0867	± 0.0276
— 8.00	.173	.125	4.18	± 0.103	± 0.0327
— 6.00	.343	.632	5.37	± 0.120	± 0.0383
— 5.19	.480	1.24	5.38	± 0.120	± 0.0384
— 5.00	.520	1.43	5.29	± 0.119	± 0.0380
— 4.59	.606	1.83	4.98	± 0.115	± 0.0366
— 4.01	.747	2.53	4.53	± 0.108	± 0.0345
— 3.01	1.05	4.23	3.84	± 0.0971	± 0.0309
— 2.50	1.17	4.95	3.62	± 0.0933	± 0.0297
— 2.00	1.36	6.08	3.29	± 0.0873	± 0.0278
— 1.94	1.39	6.26	3.24	± 0.0863	± 0.0275
— 1.00	1.67	8.02	2.88	± 0.0792	± 0.0252
0.00	1.92	9.67	2.62	± 0.0735	± 0.0234
1.00	2.14	11.1	2.42	± 0.0689	± 0.0219
2.00	2.36	12.6	2.26	± 0.0649	± 0.0206
3.00	2.53	13.9	2.17	± 0.0626	± 0.0199
4.00	2.67	14.8	2.08	± 0.0601	± 0.0191
5.00	2.77	15.6	2.03	± 0.0587	± 0.0187
6.00	2.87	16.4	1.99	± 0.0576	± 0.0183
8.00	3.04	17.6	1.90	± 0.0549	± 0.0174
10.00	3.12	18.1	1.86	± 0.0537	± 0.0171

(1) An alternative initial condition of $p_{11}(0) = 4\, p_{11}$ was used in Appendix B, where it was discovered that there is a significant dependence between the initial condition and the effective time constant of the loop. Thus the time required to reach steady-state from the large initial condition is frequently too long for an effective Monte-Carlo analysis.

verifies that in all cases the mean-squared error of the actual loop can be no lower than p_{11}. If we are interested in modulo-2 π errors we may convert the lower bound into mean-modulo-2 π-squared error by the equation :

$$\sigma_m^2 = \int_{-\infty}^{\infty} [(e + \pi) \bmod 2\pi - \pi]^2 \exp\left[\frac{-e^2}{2p_{11}}\right] de$$

$$= \int_{-\pi}^{\pi} e^2 \sum_{k=-\infty}^{\infty} \exp\left[\frac{-(e + 2k\pi)^2}{2p_{11}}\right] de, \qquad \text{(A-1)}$$

where $\sigma_m^2 \triangleq$ mean-modulo-2 π squared error. The result σ_m^2 of (A-1) is plotted in fig. A-1 as well as the equivalent discrete filter result $\sigma_{d_m}{}^2$ (based on $P_{d_{11}}$). σ_m^2 was determined by the Newton integration based on 10,000 points in the interval $[-\pi, \pi]$ and 15 standard deviations taken from the infinite sum.

The deviation the Monte-Carlo performance of the phase-locked loop from the ideal (linear) analysis can also be seen in figure A-1 to be insignificant below $N/S = -10$ db or above $+ 8$ db. The significant departure in the middle region is a result of the " cycle-slip " phenomena, whereby the unmodulated phase-error density begins to contain substantial probability in the secondary modes [8]. For extremely high noise situations, however, modulation of the error density causes the asymptotic error density to become uniform on $[-\pi, \pi]$, resulting in an asymptotic variance of $\pi^2/3 = 3.29$ (or 5.19 dB) for N/S large.

The point at which the — 10 dB departure between the two curves occurs is often referred to as " threshold ", where unlock of the loop begins to cause problems. Almost all engineering modifications to the basic loop are designed for the purpose of extending the threshold. It is clear that all nonlinear filters must have modulo-2 π error variance in the region between the two curves (discrete phase-locked loop and discrete-ideal). Thus the problem of threshold extension is equivalent to reducing steady-state error variance, which attains the maximum potential improvement of about 4 dB at — 1.8 dB N/S.

The confidence in the Monte-Carlos of the phase-locked loop can be computed using the results from [8]. Accordingly, we observed that a value for $\rho_2 = \mu_4/\mu_2^2$ must be determined, since the three-standard deviation confidence bounds on variance depend on ρ_2. Thus our Monte Carlo simulations involved the calculation of an estimate $\hat{\mu}_4$ of μ_4 and the estimate $\hat{\mu}_2$ of μ_2. Then we plotted the ratio $\hat{\mu}_4/\hat{\mu}_2^2 = \hat{\rho}_2$ in figure A-2. From the figure we determine that a good upper bound on ρ_2 is given by 5.4, which is achieved near — 5.0 dB N/S. A summary of the estimate 3 σ confidence bands is given in Table A-1, where we observe that the maximum confidence band width is from $\pm$ 0.038 dB to $\pm$ 0.120 dB, depending on whether N was taken as 500,000 (the actual number of errors averaged) or 50,000 (an approximation to the

equivalent number of independent errors used). Thus, the Monte-Carlo experiment of the phase-locked loop is more than good enough for a reliable bench-mark performance.

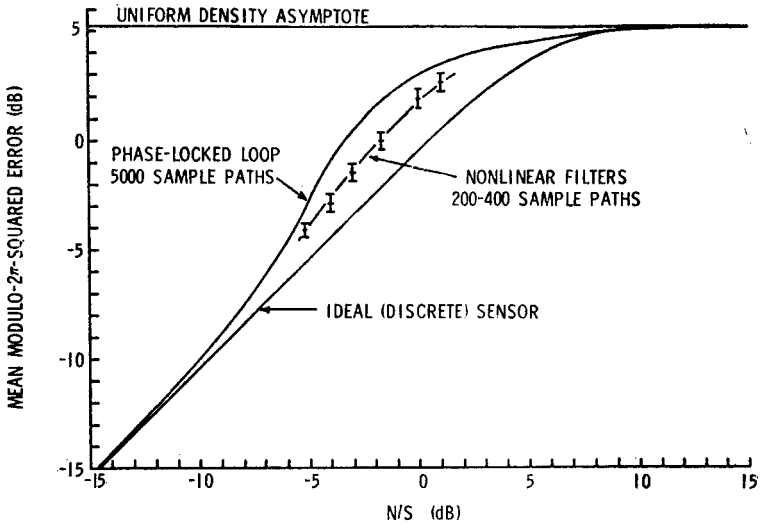


Figure A-1
MSE Performance Summary

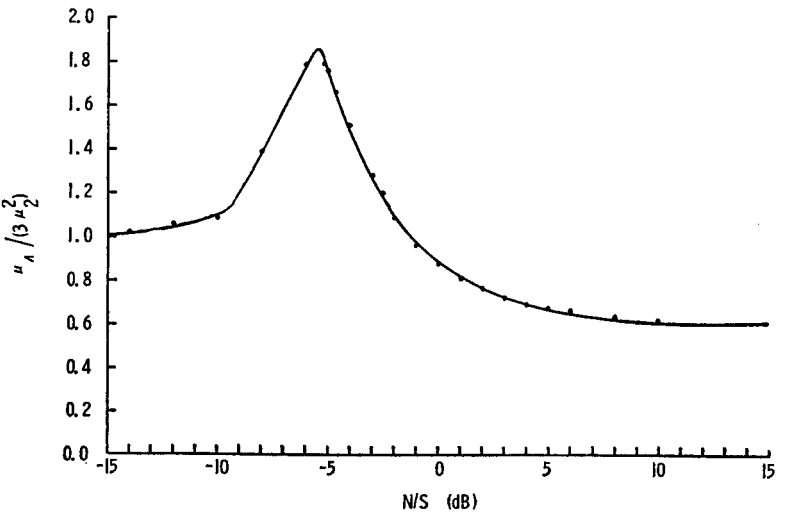


Figure A-2
Fourth Moment Divided by Three Times the Squared Variance for the Phase-Locked Loop Error

Appendix B. Numerical Experiments With The Hermite Polynomial Expansion

The objective of the numerical methods investigated in this research were to :

1) devise methods to realistically (small computation time per estimate) solve the nonlinear filter problem using contemporary computers, and

2) to demonstrate that a practical nonlinear filter can be constructed to take advantage of its inherent accuracy as compared to a linear filter. Both objectives were achieved using the methods of the previous sections and demonstrated with the programs given (Hecht [24]).

The initial numerical data given is that generated at the University of Southern California, using an IBM 360-65 computer, and reported in [24]. The subsequent data was generated at Kirtland AFB, New Mexico, using a CDC 6600 computer and is reported here for the first time.

The first goal, reduced computation time, was demonstrated by comparison to Bucy and Senne [14], where a two-dimension problem, roughly equivalent to the phase-lock problem, was solved. In the Bucy and Senne paper sophisticated techniques were used to reduce the number of computations per estimate, reducing the computation time by a factor of 200. The reduced number of computations, per estimate, was approximately 13×10^3. For the filter using Hermite expansions, described in this paper, the number of computations per estimate was 18×10^2. However, the critical calculation, the evaluation of an exponential function, had to be done only 100 times per estimate. It would be indicated, therefore, that a time improvement of a factor of 130 could be expected. The Bucy and Senne paper gave results using a Burroughs B5500 computer, and had a measured time per estimate of 45 seconds. The Hermite numerical results, given in this chapter, were obtained using an IBM 360 Model 65 computer. For the Monte-Carlo experiments described in this Appendix there were 203 sample functions, each consisting of 130 points, which ran in 120 mn, or approximately 0.273 seconds per estimate. Assuming the Burroughs B5500 was approximately equivalent to the IBM 360-65, there was a measured improvement of 165 times. The measured time on the IBM 360-65 was just about what was predicted in the referenced paper for a parallel processing computer, if such a computer should become available at a future date.

The second goal, simulation and demonstration of accuracy, was accomplished by comparing the nonlinear filter, described in Section C, with the relinearized filter described in Section B. The Monte-Carlo results obtained were possible only because of the efficiency of the Hermite method. In the following descriptive material the filter of Section B is referred to as the " linear " filter, and that of Section C as the " nonlinear " filter. Both the linear and the nonlinear filters were designed to estimate the phase angle for the phase coherent communications problem.

As indicated in the discussion of Section B, the estimate of the phase angle is required modulo 2 π. The construction of both filters was such as to attempt to track the absolute phase angle. When evaluating the filters, however, the error, modulo 2 π, was the value used.

The parameter,

$$p_{11} = E\,[(x_1 - \hat{x}_1)^2 \mid Z_\gamma, \gamma = -\infty, t]$$

was the independent parameter for all comparisons. The variances of the message-model and the observation noise were related to p_{11} as was shown in Section B. Selections of the numerical values for the initial experiments is given in Table B-1. Viterbi [42] shows the parameter $p_{11}(0)$ is also the inverse of the effective signal to noise ratio, N/S. Thus, $p_{11}(0)$ had the two physical interpretations :

1) equilibrium error variance,

3) effective noise to signal ratio.

A preliminary test was made to verify that both the linear and nonlinear filters were working properly. For small p_{11} one would expect both filters to give equal results. A sample function of 130 points (13 filter time constants) was generated for $p_{11}^{1/2} = .1$, .01, and .004; the computer listings were given in Hecht [23]. The following results were noted :

1) The sequence of estimates for the linear and nonlinear filter agree with each other to about :

$$10^{-2} \text{ radian for } p_{11}^{1/2} = .1$$
$$10^{-4} \text{ radian for } p_{11}^{1/2} = .01$$
$$10^{-5} \text{ radian for } p_{11}^{1/2} = .004$$

2) The measured variances and errors agree to within the same precision as the estimates.

3) The equilibrium computed and measured variances for the two filters agree with each other and with the linear computed value using the equations of Section B.

The nonlinear filter was tested at $\bar{p}_{11}^{1/2} = .55$ ($\bar{p}_{11} = .3025$, $N/S = -5.2$ dB) where the difference between the measured and theoretical linear variance was 3.5 dB. The nonlinear filter simulation test at $N/S = 5.2$ was under the

same conditions as the linear filter, i.e., same noise sequences, but only 200 sample functions. The phase error variance for the nonlinear filter at these conditions was — 3.55 dB, or 1.45 dB better than the linear filter. This point is also shown in figure B-1.

Table B-1
Numerical Values Used for Computer Simulations

Numerical Values Used for Computer Simulations

Δ = time between samples,
F = filter time constant,
$\frac{\Delta}{F}$ = 0.1,
$\bar{P}_{11}(0)$ = equilibrium continuous linear position error variance,
$\bar{P}_{22}(0)$ = equilibrium continuous linear velocity error variance,
$E(x_1(0)) = 0$,
$E(x_2(0)) = 0$,
$E(x_1^2(0)) = 4\,\bar{P}_{11}(0)$,
$E(x_2^2(0)) = 4\,\bar{P}_{22}(0)$,
q = continuous message driving variance = 0.01.

Number of points in each dimension for Gauss-Hermite numerical integration = 10.

Highest order of terms in series expansion of density function = 5.

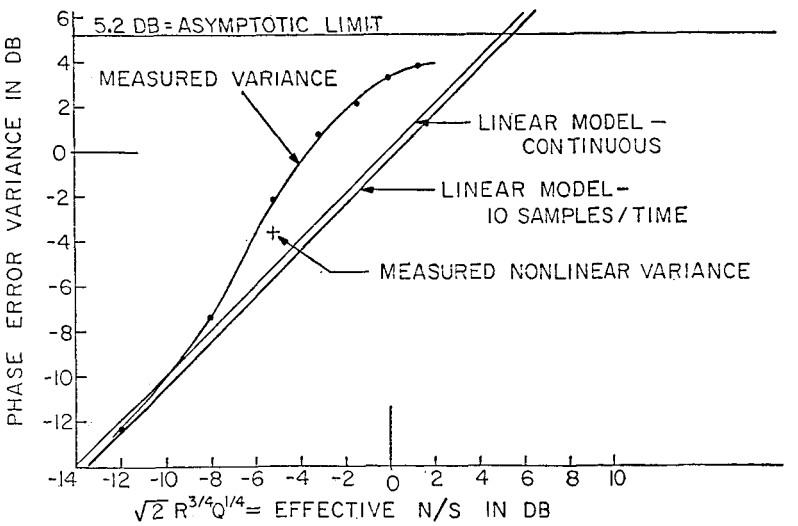


Figure B-1
Hermite Expansion Error Summary P(o) = 4 P(∞)

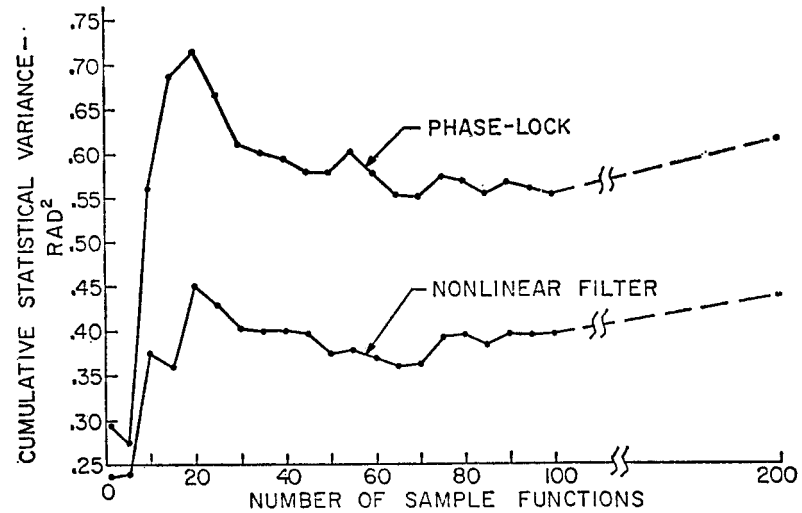


Figure B-2
Cumulative Statistical Variance P(o) = 4 P(∞)

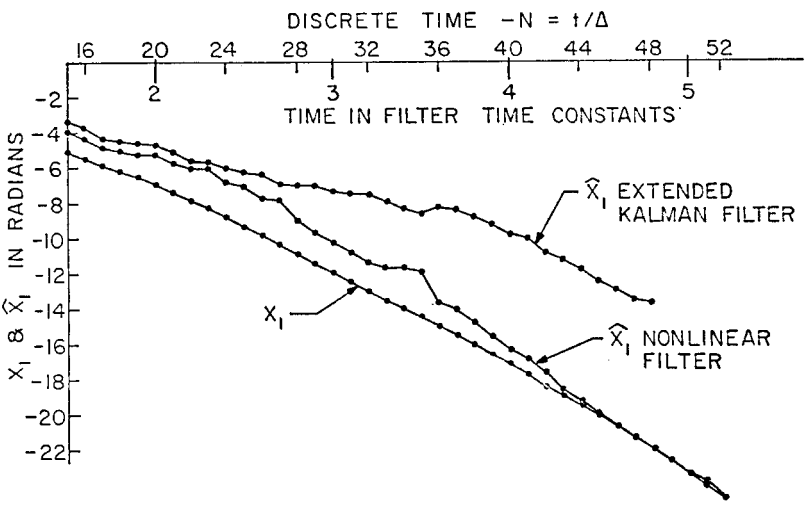


Figure B-3
Portion of Sample Function No. 6 $P_{11}(o) = 0.3025$

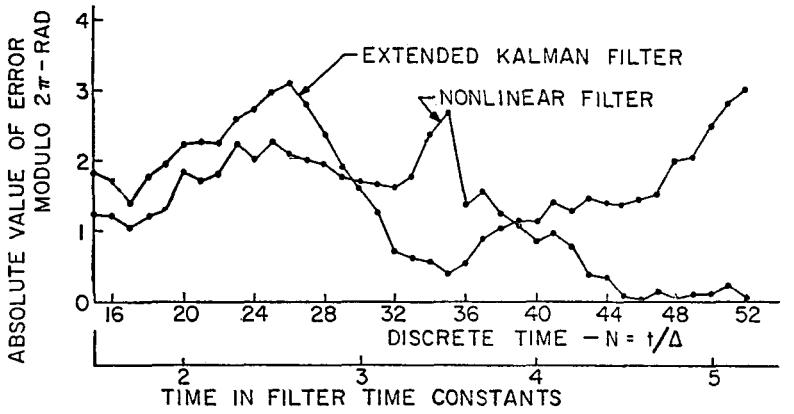


Figure B-4
Error for Sample Function No. 6 $P_{11}(o) = 0.3025$

The cumulative average variance was plotted for both filters to show the stabilization of the average as a function of number of sample functions, figure B-2. Points are plotted for every fifth sample function up until sample function number 100, then one final point at sample function number 200.

Large errors for both the linear and nonlinear filters for the phase angle problem are due to the phenomena of " cycle slippage, " which occurs more frequently as N/S gets larger. The improvement in performance of the nonlinear filter was due, primarily, to the ability of the nonlinear filter to reduce the number of cycle slips. Sample function number 6 was identified as one in which the linear filter slipped several cycles whereas the nonlinear filter held on. The sequence of estimates and errors for both filters for this sample function were analyzed and a portion of sample function number 6 (from about $N = 15$ to 52) is shown in Figure B-3 and the absolute value of the error, modulo 2 π, is plotted in figure B-4. The figures show the linear filter has slipped a cycle at $N = 35$, and is slipping a second cycle at $N = 50$. The nonlinear filter, at the same time, has a large error at $N = 35$ but appears to recover nicely and by $N = 50$ it is tracking very well.

It can be seen that the measured variance in figure B-1 is significantly different than the corresponding curve given in Appendix A (fig. A-1). Investigations showed that the initial conditions affected the variance for substantially longer than 3 time constants, as was originally assumed. The variances given in figure B-1 were based on the variance of the initial estimate being four times the equilibrium variance, whereas figure A-1 was based on the initial variance set equal to the equilibrium variance.

To demonstrate that the equilibrium solution could be achieved independent of the starting condition, one long sequence was run (83,000 points) starting at four times the equilibrium value. This is shown in figure B-5, where markers are inserted to show the result for the two curves previously mentioned for conditions given on the graph. After about 8,000 time constants (80,000 points) the cumulative average appears to be approaching the solution given in figure A-1, which was based on 5,000 Monte Carlo sample functions of 130 points each, starting at the equilibrium value.

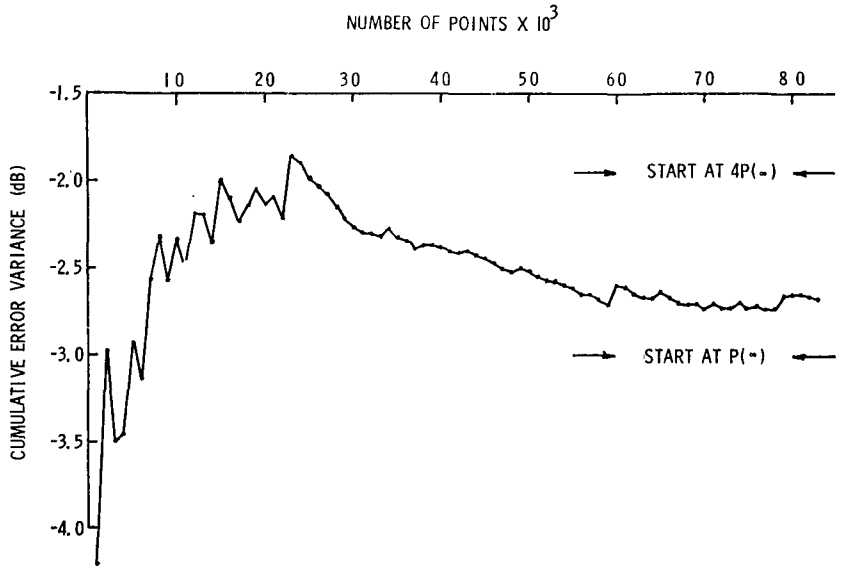


Figure B-5
Error Variance for $P_{11}(o) = -5.2$ dB Starting at $P_{11}(o) = 4\,P_{11}(\infty)$

With this new knowledge, the Hermite nonlinear filter was again evaluated with 400 Monte Carlo functions with the same parameters except that the starting values were the computed equilibrium values. This new sequence was compared to the phase-lock results for an identical set of sequences. The cumulative errors for the two filters are plotted in figure B-6, in a manner similar to figure B-2. Both the phase-lock and the nonlinear filter had smaller errors than before. The phase-lock for 400 functions was — 3.10 dB and the nonlinear filter was — 4.06 dB, for an improvement of 0.96 dB. Because of the larger number of samples the 3 σ confidence probability improved to about $\pm$ 0.07 for both filters. It was noted that the phase-lock error for 400

functions was 0.49 (— 3.10 dB) as compared to 0.50 (— 3.0 dB) for 5,000 fubctions, or within .02 (.01/.49 = .0204) of what might be considered the " correct " error.

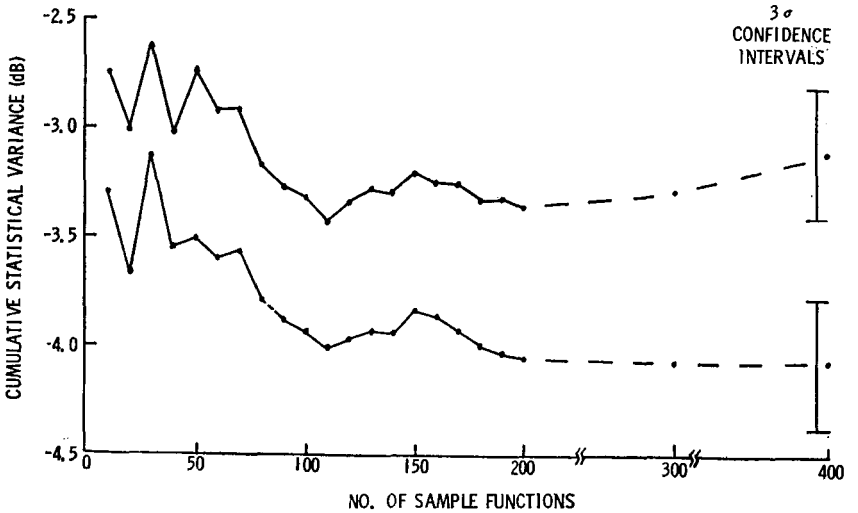


Figure B-6
Cumulative Statistical Variance for $P_{11}(o) = -5.2$ dB

The above data was generated on the CDC 6600 computer, and it was noted the nonlinear filter (Hermite expansion) computed the estimates at a rate of .127 s/estimate, which was more than two times faster than the previous computer.

In conclusion we note that a practical two-dimensional nonlinear filter was simulated using a digital computer. The digital filter error variance was within approximately 10 % of the continuous model. Using Gauss-Hermite integration and Hermite Series expansions the nonlinear filter computed solutions to a phase angle problem at the measured rate of 0.273 s per estimate on a medium speed contemporary computer and .127 s per estimate on a high speed computer, which was 165 times faster than a roughly equivalent problem using the most advanced digital techniques prior to this paper. The phase angle problem solved was a model of an existing type of communications receiver which presently uses linearization methods to handle the nonlinearities. The simulated nonlinear filter using Hermite expansions showed an error variance reduction of .96 dB at moderately high noise to signal ratio, with greater reductions at higher noise levels shown for other nonlinear filter methods. (See the discussion in the following Appendices.)

Appendix C. Cyclic Point-Mass Experiments

In this appendix we will discuss a point-mass realization of the cyclic phase density recursion. We will show how the cyclic representation of the problem is substantially better behaved for high-noise applications than the Hermite expansion, which suffers from multiple modes. We begin with a description of the special case of the cyclic point-mass filter. The density recursion satisfies the relation (see the main chapter).

$$\hat{J}_n \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = S(y_1) \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \sum_{\nu_2=-\infty}^{\infty} \exp\left\{ -\frac{1}{2\,q\Delta} \left(y_2 - \eta + \frac{2\,\pi}{\Delta} \nu_2 \right)^2 \right\} \hat{J}_{n-1} \begin{pmatrix} y_1 - \eta\Delta \\ \eta \end{pmatrix} d\eta \quad \text{(C-1)}$$

where

$$S(y_1) \triangleq C_0 \exp\left\{ -\frac{\Delta}{2\,r} \left[(z_1 - \cos y_1)^2 + (z_2 - \sin y_1)^2 \right] \right\}$$

$$= C_1 \exp\left\{ + \frac{\Delta}{r} (z_1 \cos y_1 + z_2 \sin y_1) \right\},$$

and

$$-\pi \leqslant y_1 < \pi, \quad -\frac{\pi}{\Delta} \leqslant y_2 < \frac{\pi}{\Delta}.$$

The estimate which was simulated for the cyclic densities was the cyclic estimate described in the main chapter.

For large values of $| \nu_2 |$ in the summation :

$$F(y_2, \eta) = \sum_{\nu_2=-\infty}^{\infty} \exp\left\{ -\frac{1}{2\,q\Delta} \left(y_2 - \eta + \frac{2\,\pi}{\Delta} \nu_2 \right)^2 \right\}$$

the expression is negligibly small.

Therefore, only those integers ν_2 were used where :

$$\operatorname*{Max}_{y_2, \eta} F(y_2, \eta) > \sim 10^{-20}.$$

The program that was developed makes the above test on $F(y_2, \eta)$, and in all of the results to date only the value $\nu_2 = 0$ has been found to be significant.

$F(y_2, \eta)$ is not a function of n, the integer time, and therefore was computed only one time in advance and stored for use in (C-1) for all n. A further simplification was made by taking advantage of the fact that $F(y_2, \eta)$ is only a function of $y_2 - \eta$. That is, after discretizing the argument y_2, $F(y_2, \eta)$ is computed for a range of values of $y_2 - \eta$, rather than for all combinations of y_2 and η. In the sequel we let the running variable η be called x_2^*.

The 2-dimensional interval :

$$-\pi \leqslant x_1 < \pi,$$

and

$$-\frac{\pi}{\Delta} \leqslant x_2 < \frac{\pi}{\Delta}$$

is divided into m and n equally spaced sub-intervals, respectively, and each sub-interval is defined by a point on its center, y_{1_i}, y_{2_j}, with $i = 1, ..., m$ and $j = 1, ..., n$. The points y_{1_i} and y_{2_j} define a grid which remains constant with respect to time, and the density function is represented by point masses defined only on this grid, where the magnitude of the point masses approximates the density at that point. From (C-1), the points $x^*_{2_j} = y_{2_j}$ $(j = 1, n)$; that is, for each integer j, the two grid points are identical.

The integrand in (C-1) needs to be evaluated only for those x_{2j} where Max $F(y_{2i}, x_{2j}) >$ approximately 10^{-20}. For computing an updated density $^{y_{2_i}-x^*_{2_j}}$ function $J_n \begin{pmatrix} y_{1_i} \\ y_{2_j} \end{pmatrix}$ the integrand needs to be evaluated only a small number of times for each y_{2_j} (typically 10 times for $n = 200$).

The main difficulty in the mechanization of (C-1) is in determining $J_{n-1} \begin{pmatrix} y_{1_i} - x^*_{2_j}\Delta \\ x^*_{2_i} \end{pmatrix}$, having the prior value of this function available only at a discrete set of in points in the first argument, different than those defined by $y_{1_i} - x^*_{2_j}\Delta$.

One approach to solving this problem is as follows. Let :

$$x_{1_i} = -\pi + 2\pi\frac{(i-1)}{m} + \frac{1}{2}\left(\frac{2\pi}{m}\right) \qquad i = 1, m$$

$$x^*_{2_j} = -\frac{\pi}{\Delta} + \frac{2\pi}{\Delta}\frac{(j-1)}{n} + \frac{1}{2}\left(\frac{2\pi/\Delta}{n}\right) \qquad j = 1, n \tag{C-2}$$

as defined above.

Now,

$$\begin{aligned} y_{1_i} - x^*_{2_j}\Delta &= -\pi + 2\pi\frac{(i-1)}{m} + \frac{1}{2}\left(\frac{2\pi}{m}\right) \\ &\quad - \Delta\left[-\frac{\pi}{\Delta} + \frac{2\pi}{\Delta}\frac{(j-1)}{n} + \frac{1}{2}\left(\frac{2\pi/\Delta}{n}\right)\right] \\ &= 2\pi\left(\frac{i-1}{m} - \frac{j-1}{n}\right) + \frac{1}{2}\left(\frac{2\pi}{m} - \frac{2\pi}{n}\right) \\ &= 2\pi\left(\frac{i-1}{m} - \frac{j-1}{n}\right) + \pi\left(\frac{1}{m} - \frac{1}{n}\right) \end{aligned} \tag{C-3}$$

We want to use the integers i and j to define a new grid point k, on the x_1 axis; the k grid point must agree with an original x_{1_i} grid point. That is, from (8) and (9) we want :

$$x_{1_k} = y_{1_i} - x^*_{2_j}\Delta$$

or

$$-\pi + 2\pi\left(\frac{k-1}{m}\right) - \frac{1}{2}\left(\frac{2\pi}{m}\right) = 2\pi\left[\left(\frac{i-1}{m}\right) - \left(\frac{j-1}{n}\right)\right] + \pi\left(\frac{1}{m} - \frac{1}{n}\right) \quad \text{(C-4)}$$

from which :

$$k = i - \frac{m}{n}j + \frac{1}{2}\left(\frac{m}{n} + m\right) \quad \text{(C-5)}$$

In the initial approach to this problem, when k as computed from (C-5) was not an integer, the nearest integer value was selected. In order to assure an x_1 grid point falling exactly on $x_1 = 0$ m must be an odd integer. It is also desirable to have k be correct (an integer value from (C-5)) when the x^*_2 grid point j is in the center of its range, requiring n to be an odd number. To meet the above requirements, and to subdivide such that other points, k, might match exactly with some i, the ratio $\frac{n}{m}$ should be an odd integer, which also assures that no k point will fall exactly in the middle of two adjacent i points.

A modification was made to the formula for computing k, to account for the possibility of k, as determined from (C-5), not falling in the range of (1, m).

For

$$\begin{aligned} k > m \quad & k^* = k - m \\ k < 1 \quad & k^* = k + m \end{aligned} \quad \text{(C-6)}$$

Equation (C-6) is equivalent to folding back on the x_1 axis to remain within the interval $[-\pi, \pi)$.

The above technique works satisfactorily but requires a large number of grid points to stabilize on the " true " nonlinear estimate, necessitating substantial computer expense. To test for convergence, a fixed noise sequence was generated and the filter was used to estimate the state with the number of grid points progressively increased until increasing the number of points caused no change in the sequence of estimates. It was subsequently learned that the number of points needed for convergence varied as a function of the N/S ratio. Substantial data was generated with the above method. Figure C-1 shows the phase error variance as the upper of the two curves, based on Monte-Carlos of 200 independent sample paths of 100 steady state points. The 3 σ confidence intervals represent 2,000 points (one each time constant).

An improvement which significantly reduced the computational requirements was to let k, from (11), take on non-integer values. The value of the density function at these places was evaluated by linear interpolation between the two adjacent integer values of k, where the density function was available from the prior cycle. The interpolation was required only in the x_1 direction, since the densities in the x_2 coordinate direction are computed at exactly the places where they are needed for the recursion formula. For interpolating between $k = m$ and $k = 1$, the points m and 1 were considered adjacent, completing a circle. The convergence test described above was applied to the modified filter and stabilized for a substantially smaller number of grid points. The lower curve of figure C-1 was generated using the filter with interpolation.

The two curves show the same filter error variance at about $N/S = 0$ dB, which was in the vicinity of the N/S ratio where the convergence tests were made prior to introducing the interpolation. For lower N/S the interpolating filter shows lower errors, being about one dB, less at about $N/S = -4$ dB. The numerical granularity associated with non-interpolation apparently causes significant errors at the lower N/S ratios, due to the sharpness of the phase error density function. At higher N/S ratios, the filter error density is so diffuse that the numerical errors are of no consequence.

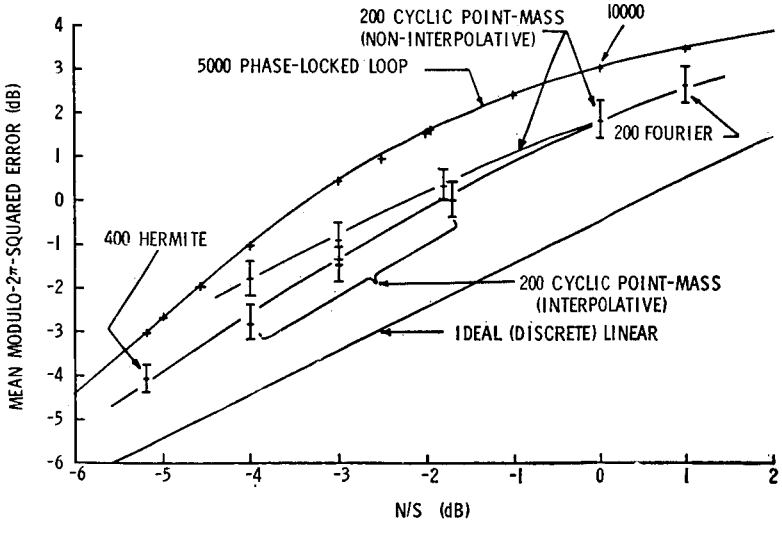


Figure C-1
Nonlinear Filter Summary (Enlarged)

The numerical results of the Monte-Carlo experiments for the cyclic point-mass filter are given in Table C-1. Table C-2 gives the associated improvement of the cyclic filter over the phase-locked loop results reported in Appendix A,

and Table C-3 gives the difference between the nonlinear filters and the ideal linear analysis. In all cases the minus and plus 3 σ confidence intervals are given, where for 2,000 points the lower threshold is — 0.394 dB below nominal and the upper threshold is 0.433 dB above the nominal.

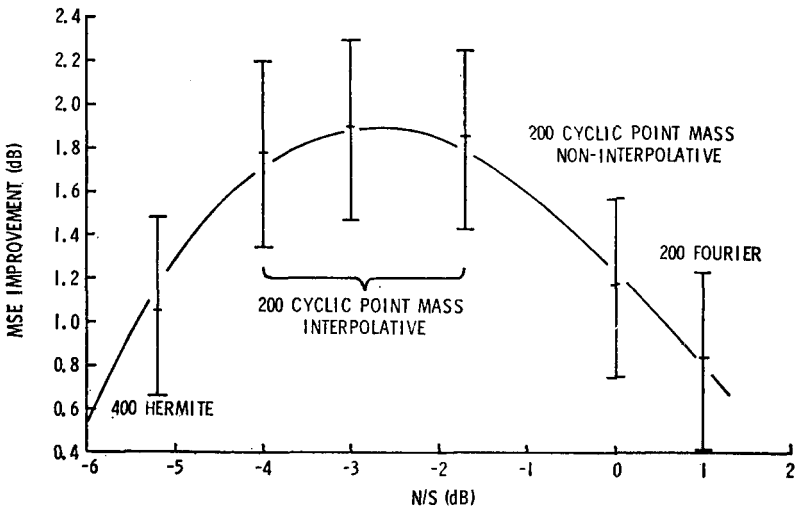


Figure C-2
MSE Improvement of Nonlinear Filters over Phase-Locked Loop

Table C-1
Monte Carlo-Mod 2π Error Performance Data for the Cyclic Point Mass Estimates

N/S (dB)	Non-Interpolative MSE (dB) (Mod 2π)			Interpolative MSE (dB) (Mod 2π)		
	Low	Nom.	High	Low	Nom.	High
— 4.01	— 2.20	— 1.81	— 1.38	— 3.22	— 2.83	— 2.40
— 3.01	— 1.33	— 0.94	— 0.51	— 1.87	— 1.48	— 1.05
— 1.79	— 0.09	0.30	0.73	—	—	—
— 1.70	—	—	—	— 0.38	0.01	0.44
0.00	1.46	1.85	2.28	—	—	—

Table C-2
Monte-Carlo Improvements Cyclic Point-Mass over Phase-Locked Loop

N/S (dB)	Non-Interpolative MSE (dB) Improvement			Interpolative MSE (dB) Improvement		
	Low	Nom.	High	Low	Nom.	High
— 4.01	0.33	0.76	1.15	1.35	1.78	2.17
— 3.01	0.93	1.36	1.75	1.47	1.90	2.29
*— 1.79	1.02	1.45	1.84	—	—	—
**— 1.70	—	—	—	1.41	1.84	2.23
0.00	0.75	1.18	1.57	—	—	—

Table C-3
Monte-Carlo Difference Between Cyclic Point-Mass and Ideal Linear

N/S (dB)	Non-Interpolative MSE (dB) Difference			Interpolative MSE (dB) Difference		
	Low	Nom.	High	Low	Nom.	High
— 4.01	2.23	2.62	3.05	1.21	1.60	2.03
— 3.01	2.10	2.49	2.92	1.56	1.95	2.38
— 1.79	2.13	2.52	2.95	—	—	—
— 1.70	—	—	—	1.75	2.14	2.57
0.00	1.9	2.29	2.72	—	—	—

* Phase-Locked Performance read from graph (1.75 dB).
** Phase-Locked Performance read from graph (1.85 dB).

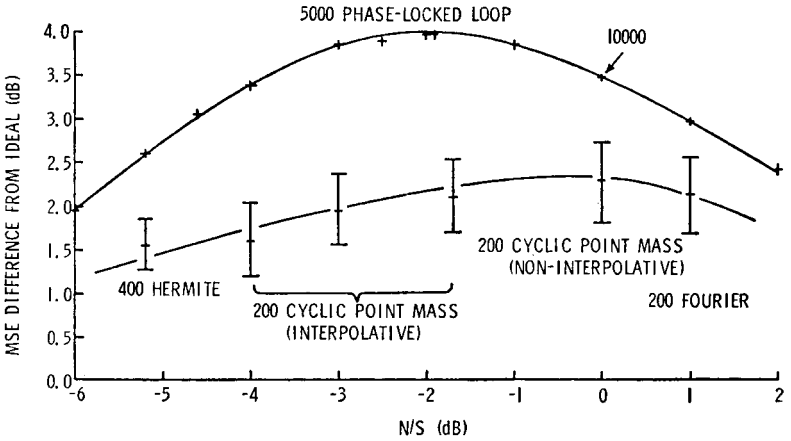


Figure C-3
MSE Difference from Ideal Linear Analysis

The summary of these data in figure C-1 is shown with the phase-locked loop performance and the Idealized Linear reference curve. Figure C-2 shows the improvements over the phase-locked loop and figure C-3 shows the difference between nonlinear and ideal. In all figures the Hermite point from Appendix B is superimposed.

In addition to determining the Monte-Carlo performance, a study was done to determine the execution times for various grid sizes in an effort to obtain a cost versus performance comparison.

The results of figure C-1 were obtained with a grid of $m = 21$ and $n = 105$. The number p is evaluated by the program for each problem condition and represents the number of points on each side of the point (x_{1_i}, x_{2_j}) in the x_2 direction, which contribue to the computation of the density of that point.

An estimate of the time required to update the density function was based on the knowledge that the time was roughly proportional to the number of computations. For each, m, n it requires $m \times n \times p$ computations. The measured data for the 21×105 grid case was about 0.215 seconds per estimate (as compared to the 5-coefficient Hermite value of 0.121 seconds per estimate). Using the above scaling, the data of Table C-4 was generated.

To determine an adequate grid size many runs were made with the same random sequence inputs, using different grid size combinations. All runs were made with $N/S = -1.7$ dB, Filter Time constant $= 5.13$ seconds, samples per time constant $= 10$, $q = .010$ and $r = 1.735$. The sequence of estimates for a ten-increment time period (integer time $= 31,40$) were compared with each other. It was desired to find a combination which gave reasonably

good results (as compared to larger grids) while minimizing the time per estimate. Table C-5 gives 3 sequences with the grid size progressively increasing while maintaining the ratio $n/m = 5$. Table C-6 maintains $n = 155$ while varying the ratio n/m from 3 through 5. Table C-7 maintains $m = 31$ while varying the ratio n/m from 5 through 7. Analyzing the absolute error sequences from Tables C-5 through C-7 we note that for $n/m = 5$, and increasing grid size, the errors become smaller as the grid size increases, with very little change between $m = 31$ and $n = 41$, and insignificant differences between $m = 21$ $n = 41$. For $n = 155$, Table C-6, and for $m = 31$, Table C-7, the same general trend was observed. That is, slight but insignificant improvement with increasing grid size. For Table C-7, especially, there appears to be no advantage in increasing m. Tables C-5 and C-6 suggest, however, that ultimate stabilization can be achieved by increasing m with n about 155. The choice of a 21 × 105 grid for the Monte Carlo experiments was based on a compromise between time and accuracy, as illustrated in the Tables.

Table C-4
Timing Estimates

Grid	p	m/n	Time/Est Sec.
21 × 105	4	5	.215
31 × 155	6	5	.700
41 × 205	8	5	1.63
31 × 93	4	3	.279
31 × 125	5	4	.468
31 × 155	6	5	.700
31 × 195	8	6	1.17
31 × 217	9	7	1.46
21 × 145	6	7	.442
25 × 151	6	6	.547
31 × 155	6	5	.700
39 × 155	6	4	.875
51 × 155	6	3	1.15

m = number of lines in x_1 direction,
n = number of lines in x_2 direction,
$2p + 1$ = number of computations to update each grid point.

Table C-5
***n/m* Constant**

$\frac{n}{m} = 5$	Time Est.	.215 Sec. $m = 21$ $n = 105$		.700 Sec. $m = 31$ $n = 155$		1.63 Sec. $m = 41$ $n = 205$	
Time	x_1	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$
31	— 2.394	3.094	.789	3.100	.789	3.106	.783
32	— 2.296	— 3.085	.789	— 3.081	.785	— 3.075	.779
33	— 2.244	— 2.146	.098	— 2.216	.028	— 2.234	.010
34	— 2.220	— 1.905	.315	— 1.943	.277	— 1.955	.265
35	— 2.210	— 2.146	.064	— 2.163	.047	— 2.172	.038
36	— 2.221	— 2.903	.682	— 2.880	.659	— 2.868	.647
37	— 2.222	— 1.908	.314	— 1.912	.310	— 1.921	.301
38	— 2.198	— 1.246	.951	— 1.274	.924	— 1.287	.911
39	— 2.162	— 1.312	.850	— 1.330	.832	— 1.343	.819
40	— 2.195	— 1.894	.301	— 1.866	.329	— 1.868	.327

Table C-6
***n* Constant**

$n = 155$	.700 Sec. $m = 31$ $\frac{n}{m} = 5$		.875 Sec. $m = 39$ $\frac{n}{m} \cong 4$		1.15 Sec. $m = 51$ $\frac{n}{m} \cong 3$	
Time	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$
31	3.100	.789	3.105	.784	3.106	.783
32	— 3.081	.785	— 3.077	.781	— 3.076	.780
33	— 2.216	.028	— 2.232	.012	— 2.246	.002
34	— 1.943	.277	— 1.954	.266	— 1.960	.260
35	— 2.163	.047	— 2.172	.038	— 2.177	.033
36	— 2.880	.659	— 2.870	.649	— 2.867	.646
37	— 1.912	.310	— 1.920	.302	— 1.925	.297
38	— 1.274	.924	— 1.287	.911	— 1.290	.908
39	— 1.330	.832	— 1.343	.829	— 1.347	.825
40	— 1.866	.329	— 1.871	.324	— 1.870	.323

Table C-7
***m* Constant**

$m = 31$	.700 Sec. $m = 155$ $\frac{n}{m} = 5$		1.17 Sec. $m = 195$ $\frac{n}{m} \cong 6$		1.46 Sec. $m = 217$ $\frac{n}{m} \cong 7$	
Time	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$	$\hat{x}_1$	$\lvert \hat{x}_1 - x_1 \rvert$
31	3.100	.789	3.100	.789	3.101	.788
32	— 3.081	.785	— 3.081	.785	— 3.081	.785
33	— 2.216	.028	— 2.216	.028	— 2.212	.024
34	— 1.943	.277	— 1.943	.277	— 1.942	.278
35	— 2.163	.047	— 2.163	.047	— 2.162	.048
36	— 2.880	.659	— 2.879	.658	— 2.878	.657
37	— 1.912	.310	— 1.912	.310	— 1.912	.310
38	— 1.274	.924	— 1.274	.924	— 1.275	.923
39	— 1.330	.832	— 1.330	.832	— 1.331	.831
40	— 1.866	.329	— 1.867	.328	— 1.869	.326

Appendix D. A Fourier Series Experiments

Mallinckrodt, Bucy, and Cheng [31] have observed the fact that since the cyclic phase density is periodic, a Fourier Series appears appropriate for representation of the density functions. They have developed equations for the evolution of the Fourier Series for the one dimensional problem. We extend their analysis to our two-dimensional problem and present the preliminary results of a numerical experiment in this appendix.

We begin by observing that an arbitrary function $J(y)$ periodic on the rectangle $-\pi \leqslant x_1 \leqslant \pi, -\frac{\pi}{\Delta} \leqslant x_2 \leqslant \frac{\pi}{\Delta}$ may be represented in terms of its two d mensional Fourier Series :

$$J_n \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \sum_m \sum_l a^n_{ml} \, e^{+imy_1} \, e^{+il\Delta y_2}, \tag{D-1}$$

where

$$a^n_{ml} = \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \int_{-\pi}^{\pi} J_n \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} e^{-imy_1} \, e^{-il\Delta y_2} \, dy_1 \, dy_2 \tag{D-2}$$

Now the cyclic density obeys the recursion relation :

$$\tilde{J}_n \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = S(y_1) \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} M(y_2 - x_2)\, \tilde{J}_{n-1} \begin{pmatrix} y_1 - x_2 \Delta \\ x_2 \end{pmatrix} \mathrm{d}x_2, \qquad \text{(D-3)}$$

where

$$S(y_1) = C_0 \exp \left\{ \frac{z_1 \cos y_1 + z_2 \sin y_1}{r/\Delta} \right\}, \qquad \text{(D-4)}$$

and

$$M(u) = \sum_k \exp \left\{ \frac{-\left(u + \dfrac{2\pi k}{\Delta}\right)^2}{2q\Delta} \right\}$$

Expressing $M(u)$ as a Fourier Series yields :

$$M(u) = \sum_\nu m_\nu \, \mathrm{e}^{i\nu\Delta u}, \qquad \text{(D-6)}$$

where

$$m_\nu = \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \sum_k \exp \left\{ \frac{-\left(u + \dfrac{2\pi k}{\Delta}\right)^2}{2q\Delta} \right\} \mathrm{e}^{-i\nu\Delta u}\, \mathrm{d}u \qquad \text{(D-7)}$$

$$= \int_{-\infty}^{\infty} \exp \left\{ -\frac{u^2}{2q\Delta} + i\nu\Delta u \right\} \mathrm{d}u$$

$$= 2\pi \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left[-\frac{u^2}{2q\Delta} \right] \exp \left[-iu(-\nu\Delta) \right] \mathrm{d}u \right\}$$

$$= 2\pi \left\{ \frac{\sqrt{q\Delta}}{\sqrt{2\pi}} \exp \left[-\frac{q\Delta}{2} (-\nu\Delta)^2 \right] \right\}$$

$$= \sqrt{2\pi q\Delta} \exp \left[-\frac{q\nu^2\Delta^3}{2} \right] \qquad \text{(D-8)}$$

Next, we represent $S(y)$ by an infinite series by making the substitution $z_1 + iz_2 = |\, z \,| \exp(i\emptyset)$, so that $z_1 = |\, z \,| \cos \emptyset$, and $z_2 = |\, z \,| \sin \emptyset$. Then $S(y)$ is expressed as :

$$S(y) = C_0 \exp \left\{ \frac{|\, z \,|}{\Delta/r} \cos (y - \emptyset) \right\}$$

$$= C_0 \sum_\lambda I_\lambda \left(\frac{|\, z \,|}{\Delta/r} \right) \exp \{ i\lambda (y - \emptyset) \},$$

$$= C_0 \sum_\lambda S_\lambda \exp \{ i\lambda y \} \qquad \text{(D-9)}$$

where I_λ is the modified Bessel function of imaginary argument of order λ (see Abramowitz and Stegun [1], Equation 9.6.34).

Finally, we combine the definition (D-2) with the expression (D-3), thereby obtaining :

$$a^n_{ml} = \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \int_{-\pi}^{\pi} S(y_1) \left[\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} M(y_2 - x_2)\, \tilde{J}_{n-1} \begin{pmatrix} y_1 - x_2\Delta \\ x_2 \end{pmatrix} dx_2 \right] e^{-imy_1}\, e^{-il\Delta y_2}\, dy_1\, dy_2$$

(let $\tau = y_2 - x_2$, or $y_2 = \tau + x_2$)

$$= \left[\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} M(\tau)\, e^{-il\Delta\tau}\, d\tau \right] \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \int_{-\pi}^{\pi} S(y_1)\, \tilde{J}_{n-1} \begin{pmatrix} y_1 - x_2\Delta \\ x_2 \end{pmatrix} e^{-imy_1}\, e^{-il\Delta x_2}\, dy_1\, dx_2$$

(identify m_l from (D-7))

$$= m_l \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \int_{-\pi}^{\pi} S(y_1)\, \tilde{J}_{n-1} \begin{pmatrix} y_1 - x_2\Delta \\ x_2 \end{pmatrix} e^{-imy_1}\, e^{-il\Delta x_2}\, dy_1\, dx_2$$

(substitute (D-1) for $\tilde{J}_{n-1}$ and (D-9) for $S(y_1)$)

$$= m_l \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \int_{-\pi}^{\pi} C_0 \sum_\lambda S_\lambda\, e^{i\lambda y_1} \sum_j \sum_k \left[a^{n-1}_{jk}\, e^{ij(y_1 - x_2\Delta)}\, e^{ik\Delta x_2} \right] \cdot e^{-imy_1}\, e^{-il\Delta x_2}\, dy_1\, dx_2$$

(rearrange using Fubini theorem)

$$= C_0\, m_l \sum_\lambda \sum_j \sum_k S_\lambda\, a^{n-1}_{jk} \int_{-\pi}^{\pi} e^{iy_1(j+\lambda-m)}\, dy_1 \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} e^{i\Delta x_2(k-l-j)}\, dx_2$$

(observe $= 0$ unless $j = m - \lambda$ and $k = j + l$)

$$= \frac{4\pi^2}{\Delta} C_0\, m_l \sum_\alpha S_{m-\alpha}\, a^{n-1}_{\alpha, l+\alpha}. \qquad \text{(D-10)}$$

From the definition (D-2) we observe that $a^n_{00} = 1$, since $\tilde{J}_n$ must have unit total integral. But :

$$a^n_{00} = \frac{4\pi^2}{\Delta} C_0\, m_0 \sum_\alpha S_{-\alpha}\, a^{n-1}_{\alpha,\alpha} = 1. \qquad \text{(D-11)}$$

Accordingly, we have

$$\frac{4\pi^2}{\Delta} C_0 = \frac{1}{m_0 \sum_\alpha S_{-\alpha}\, a^{n-1}_{\alpha,\alpha}},$$

So, if we define $\tilde{m}_\nu = m_\nu/m_0$, we have finally that

$$a^n_{ml} = \frac{\tilde{m}_l \sum_\alpha S_{m-\alpha}\, a^{n-1}_{\alpha,\, l+\alpha}}{\sum_\alpha S_{-\alpha}\, a^{n-1}_{\alpha,\,\alpha}}, \tag{D-12}$$

where

$$\tilde{m}_l = \exp\left[-\frac{ql^2\Delta^3}{2}\right], \tag{D-13}$$

and

$$S_\alpha = I_\alpha\left(\frac{|z|}{\Delta/r}\right)\exp\{-i\alpha\emptyset\}. \tag{D-14}$$

Next, we observe that (from (D-2) :

$$a^n_{-1,\,0} = \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}}\int_{-\pi}^{\pi} e^{+iy_1}\, J_n\begin{pmatrix} y_1 \\ y_2\end{pmatrix} dy_1\, dy_2 = E\,[\cos x_1 \mid Z_n] + iE\,[\sin x_1 \mid Z_n]. \tag{D-15}$$

Therefore, the cyclic estimate (which minimizes $E\,[2\,(1 - \cos e)]$) is given by :

$$\hat{x}_1(n \mid n) = \tan^{-1}\{E\,[\sin x_1 \mid Z_n]/E\,[\cos x_1 \mid Z_n]\}$$
$$= \tan^{-1}\{Im\,(a^n_{-1,\,0})/Re\,(a^n_{-1,\,0})\} \tag{D-16}$$

Using the equations (D-12) — (D-14) and (D-16) we have implemented an example of the Fourier Series filter with $-5 \leqslant m \leqslant 5$ and $-5 \leqslant l \leqslant 5$ for a total of $11 \times 11 = 121$ coefficients. Due to the occurrences of negative mass we discovered that Fourier Series is not suitable for low-noise situations. On the other hand, for $N/S = 1$ dB and $q = 0.1$ we managed to get quite good results. This may be in figure C-1, where the result of the mean-modulo-2 π-squared error of the Fourier Series is shown in conjunction with the experimental results for the phase-locked loop (Appendix A), the Hermite Expansion (Appendix B), and the point-mass representation (Appendix C). The nominal Monte Carlo result for the Fourier Series at $N/S = 1$ dB was 2.65 dB with a 200 Monte-Carlo 3 σ confidence from 2.26 dB to 3.08 dB. This result is equivalent to a nominal improvement over the phase-locked loop of 0.84 dB with 3 σ confidence from 0.41 dB to 1.23 dB. Also, the difference between Fourier Series performance and the ideal linear was nominally 2.12 dB with confidence interval from 1.73 dB to 2.55 dB.

Although the above Monte-Carlo result was consistent with the previous experiments it represents a preliminary result, since we still have not completely isolated a solution to the negative mass dilemma. More results will follow in a later paper.

Appendix E. A Movie of Conditional Densities

Just as we have demonstrated the value of visual inspection of conditional densities for the tracking problem in the past see [14], we now illustrate the wealth of information contained in the conditional densities for the phase demodulator. In this appendix, we describe a movie made of cyclic phase densities using the point-mass method described in Appendix C.

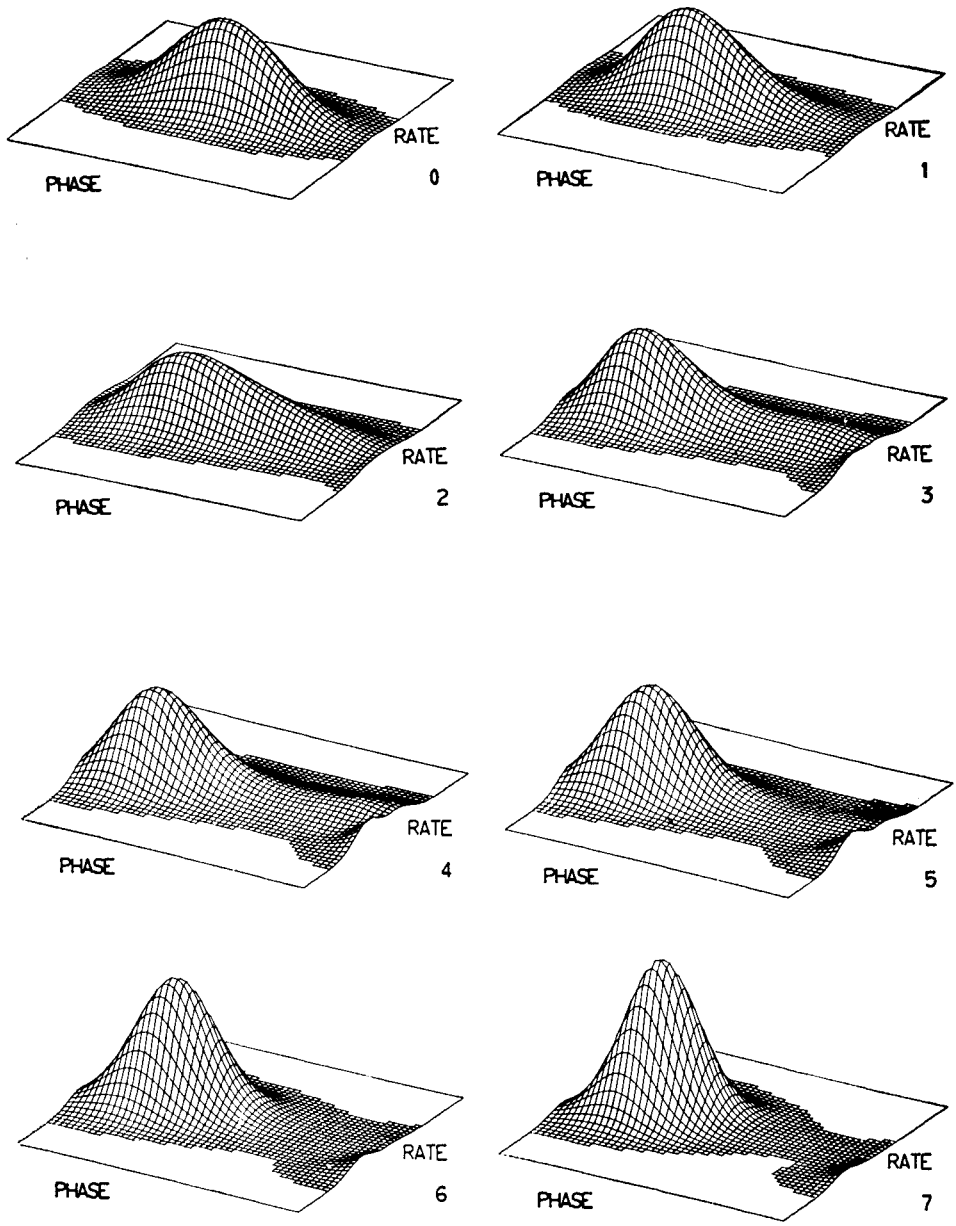
The parameters used for the sequence in the movie were as follows :

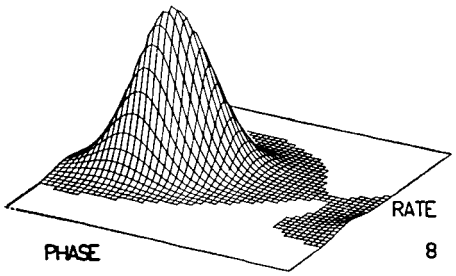
$$N/S = 0 \text{ dB}, \qquad q = 0.1$$
$$\Delta = 0.24$$

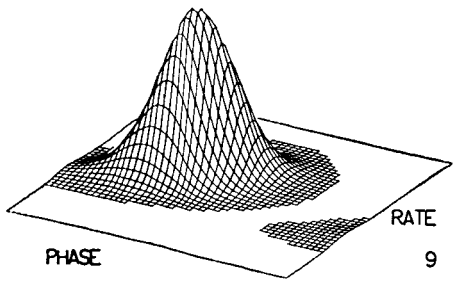
The initial density was chosen with the nominal steady-state value $P_{11} = 1.85$ dB (1.53), and the grid size was set at 31 points in phase by 155 points in phase-rate. The isometric views of the densities, seen in figure E-1, are shown for phase over the entire interval $[-\pi, \pi]$, but phase rate is shown over only one third of the interval $\left[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}\right]$. Thus spillover in the phase rate direction is not lost, but merely not shown. In the initial sequence from the movie [11], shown in figure E-1, many features may be observed. Cycle slips in phase are accompanied by general turbulence of the density, the appearance of multiple modes and other anomalies. One explanation for the appearance of multiple modes and thus the cycle slips is the occasional major disagreement between the in-line and quadrature measurement components z_1 and z_2, as a result of the independent noises v_1 and v_2. The sequence shown in the figure illustrates, though, how recovery is gradually reaccomplished when the measurements agree.

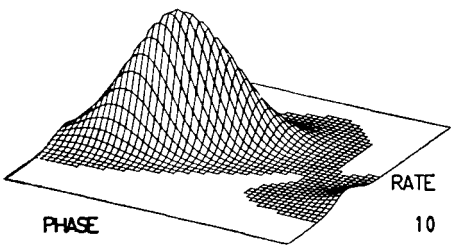
The following 18 pages constitute figure E-1. The densities are taken from the initial condition and 59 conditional a posteriori densities.

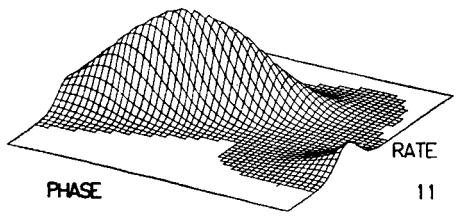
Figure E-1 (A à O)
A Typical Sample Path of Densities Evolving in Time

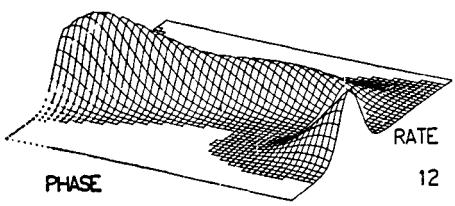


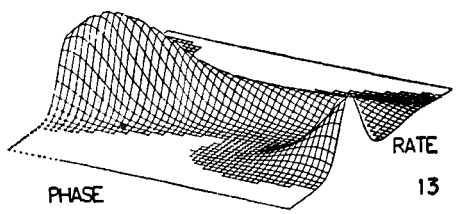

RATE
PHASE
8

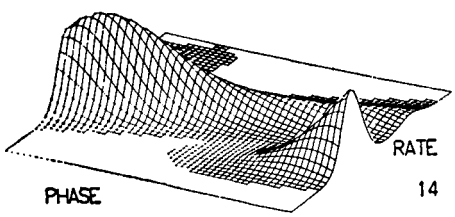

RATE
PHASE
9

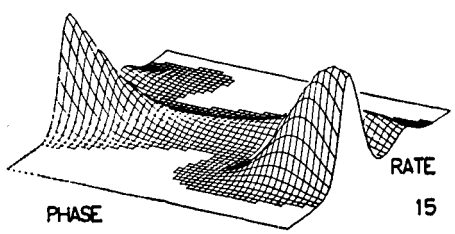

RATE
PHASE
10

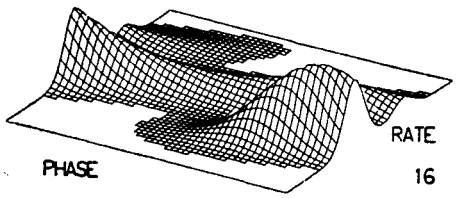

RATE
PHASE
11

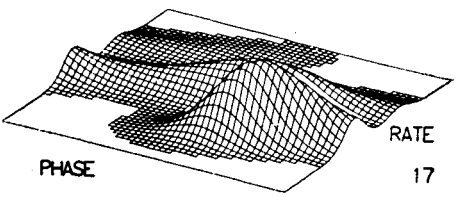

RATE
PHASE
12

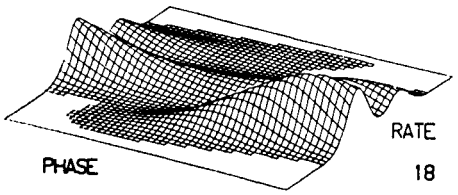

RATE
PHASE
13

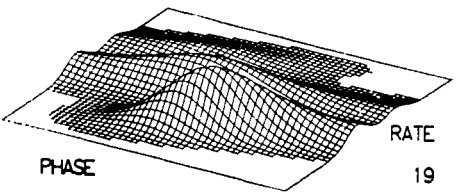

RATE
PHASE
14

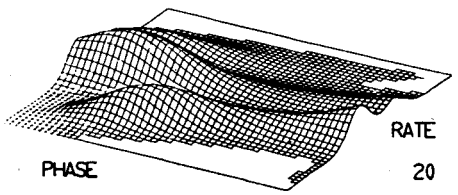

RATE
PHASE
15

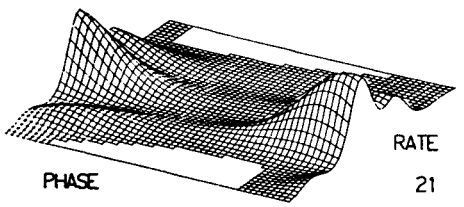

RATE
PHASE
16

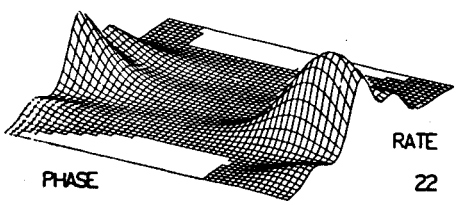

RATE
PHASE
17

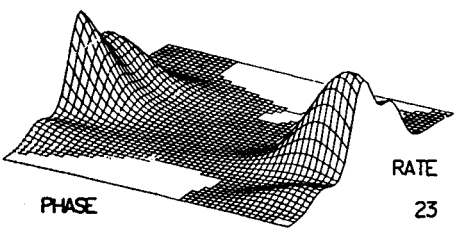

RATE
PHASE
18

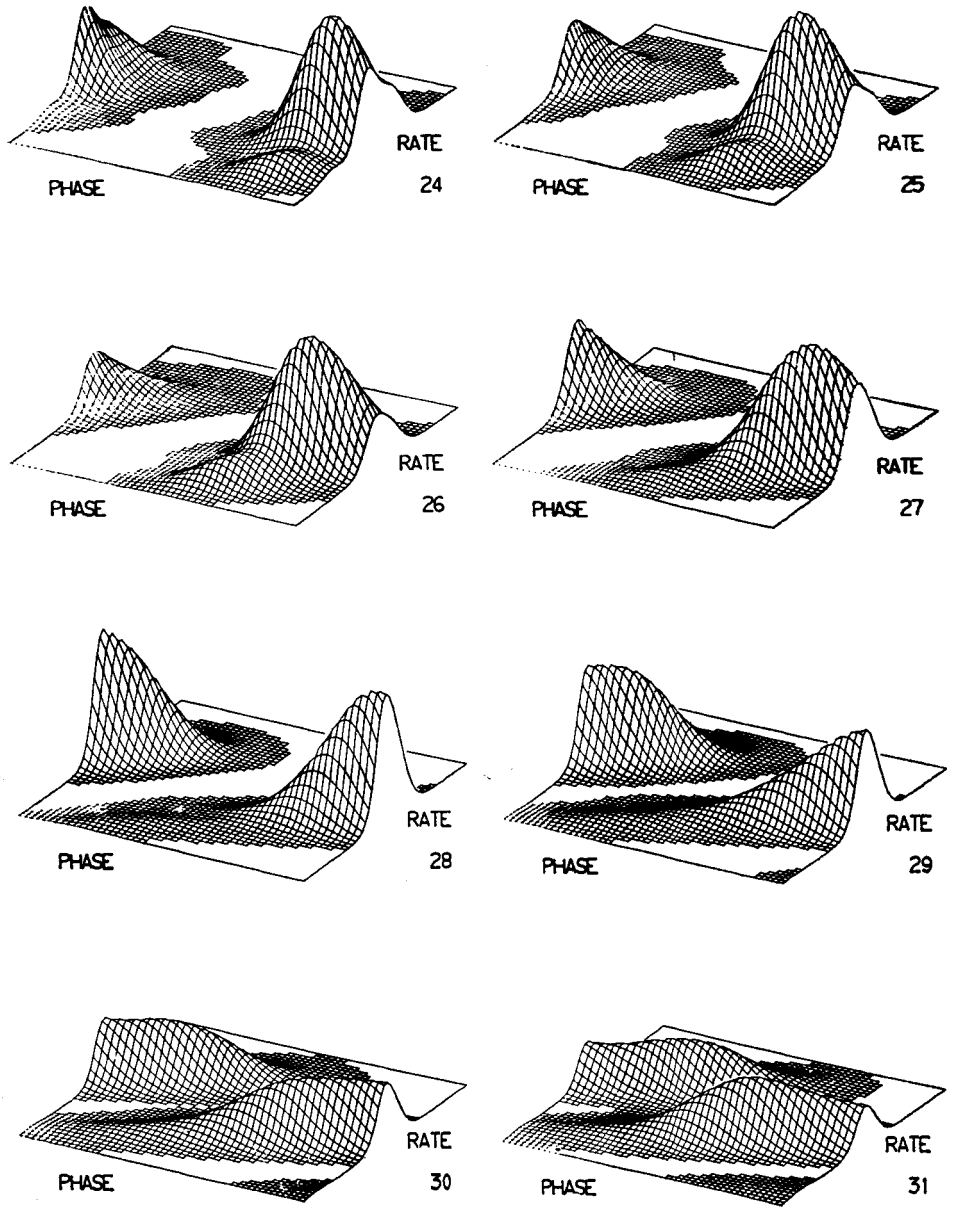

RATE
PHASE
19

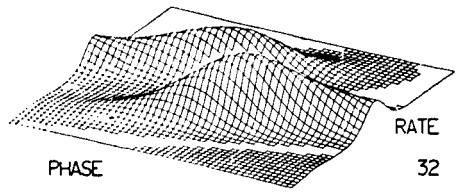

RATE
PHASE
20

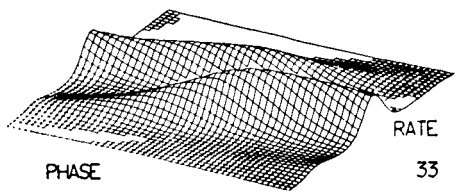

RATE
PHASE
21

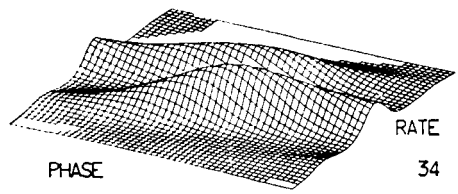

RATE
PHASE
22

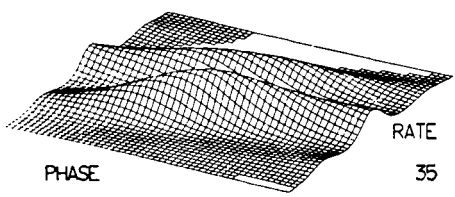

RATE
PHASE
23

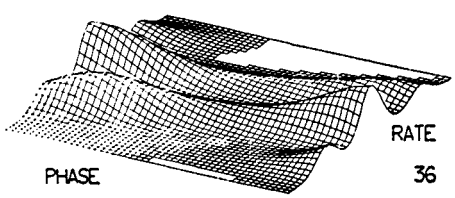

PHASE
RATE
24
PHASE
RATE
25
PHASE
RATE
26
PHASE
RATE
27
PHASE
RATE
28
PHASE
RATE
29
PHASE
RATE
30
PHASE
RATE
31

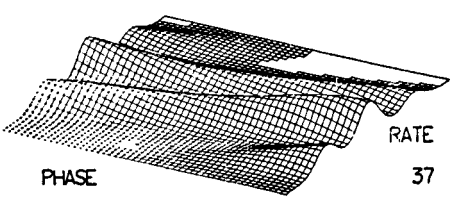

RATE
PHASE
32

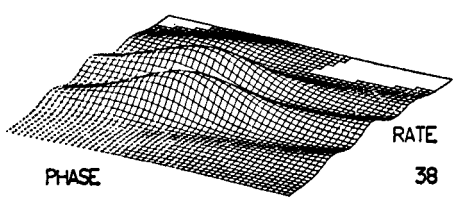

RATE
PHASE
33

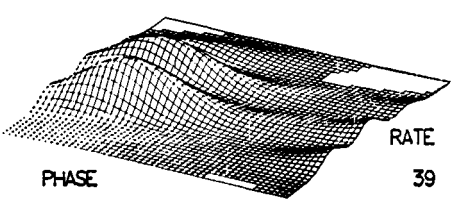

RATE
PHASE
34

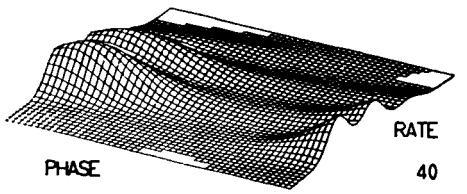

RATE
PHASE
35

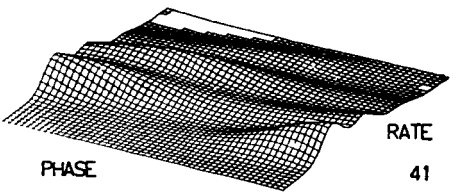

RATE
PHASE
36

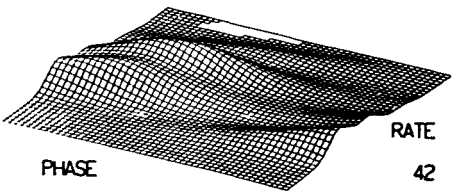

RATE
PHASE
37

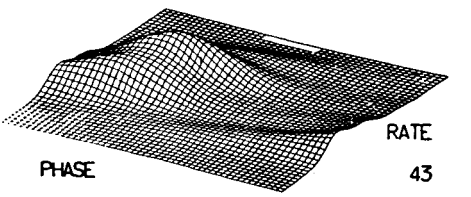

RATE
PHASE
38

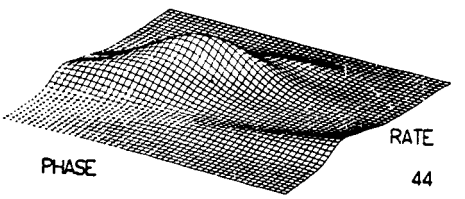

RATE
PHASE
39

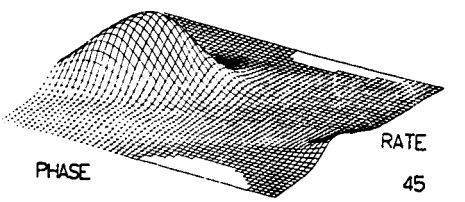

RATE
PHASE
40

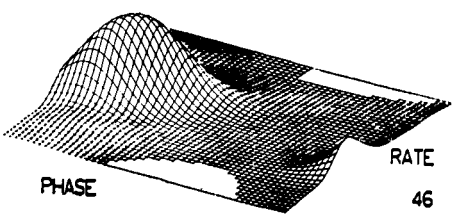

RATE
PHASE
41

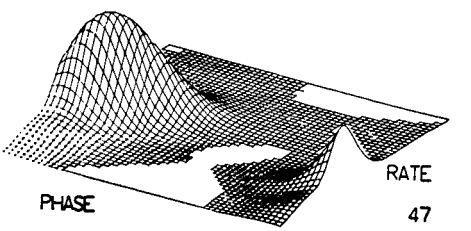

RATE
PHASE
42

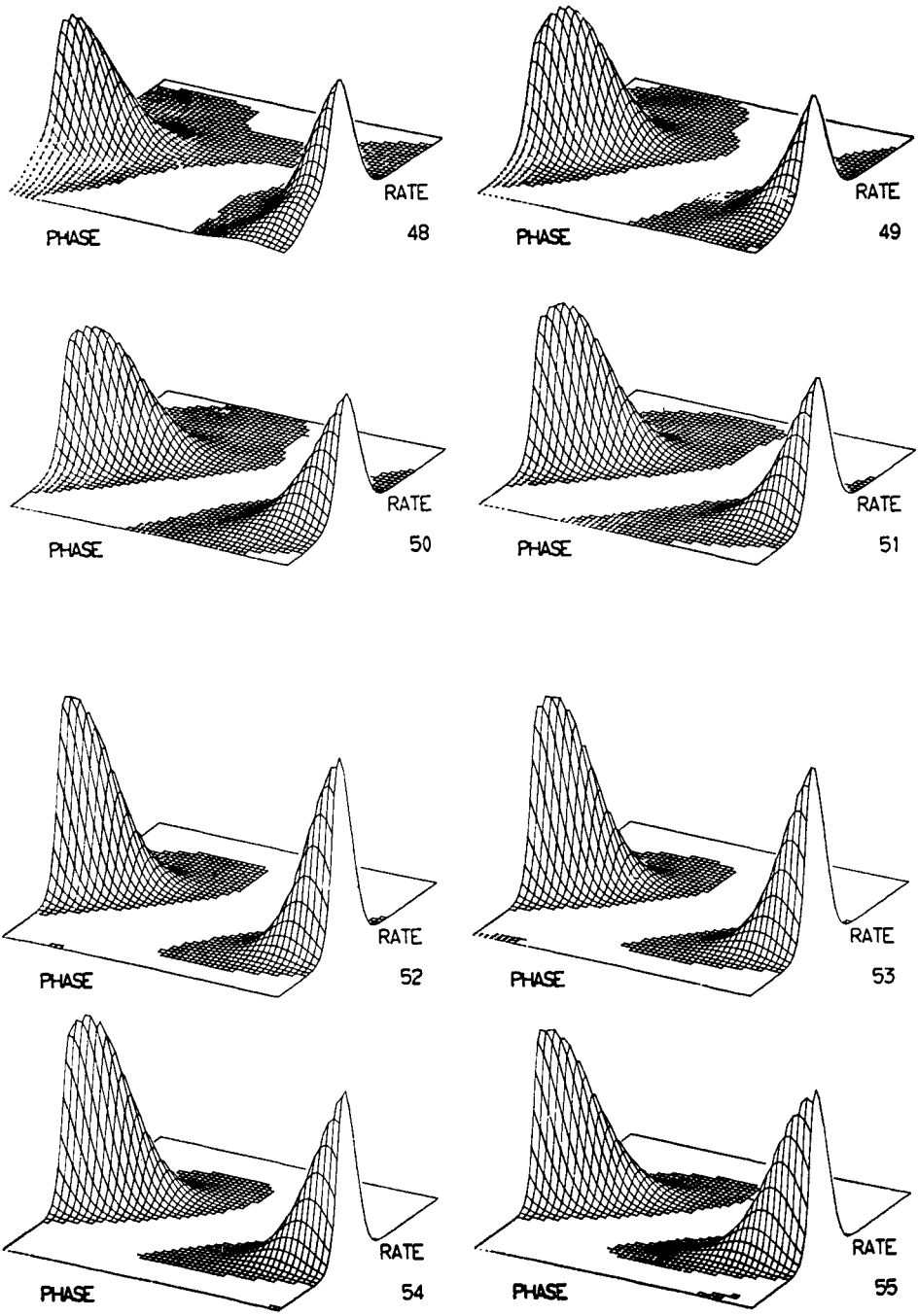

RATE
PHASE
43

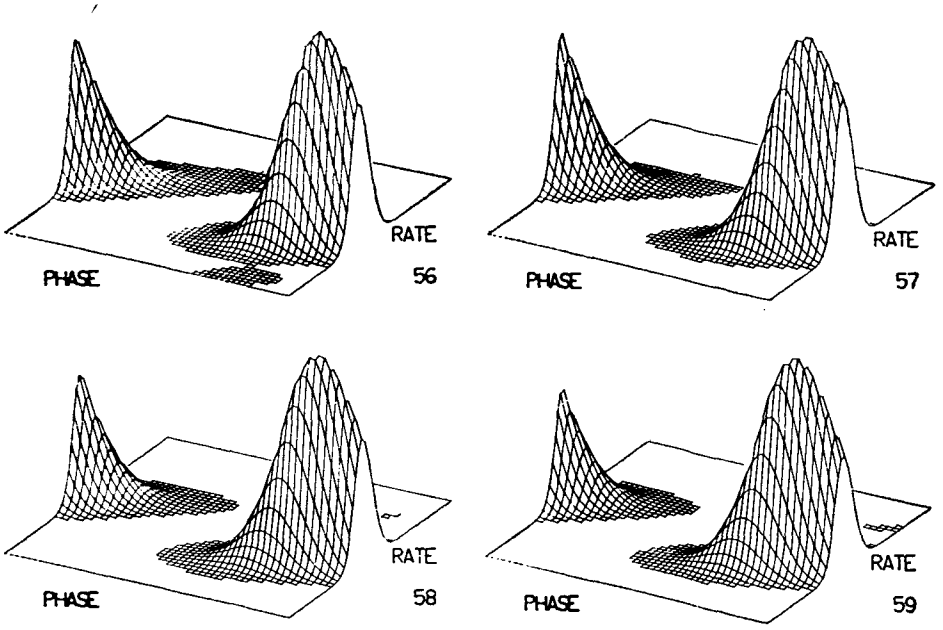

RATE
PHASE
44


RATE
PHASE
45


RATE
PHASE
46


RATE
PHASE
47


PHASE
RATE
48
PHASE
RATE
49
PHASE
RATE
50
PHASE
RATE
51
PHASE
RATE
52
PHASE
RATE
53
PHASE
RATE
54
PHASE
RATE
55



CONCLUSION

We have shown how it is possible to design and build an optimal phase demodulation system which achieves considerable mean-square error performance improvement over the filter from the classical phase-locked loop. In order to build and evaluate this optimal filter, we have employed a CDC 6600 computer. Of course, only the most important problems would merit the allocation and dedication of such an expensive tool to phase tracking. We would envision the development of a special purpose computer exploiting the innate parallelism of the Bayes rule calculation, necessary to find the requisite conditional density, to realize the optimal filter. In fact in [13] and [34] we demonstrated the speed and effectiveness of such a device, based on a hybrid computer. Abstractly, Bayes' rule may be written as :

$$J_{n+1/n} = S_* J_{n/n}$$

and

$$J_{n/n} = D \cdot J_{n/n-1}$$

with * denoting spatial convolution and · denoting pointwise multiplication. Rewriting the equations in this form, suggests that the convolution equation could be realized by a fast Fourier Transform hardware box, in conjunction with a mini-computer to close the loop. The computational burden is the

convolution equation, and for the two-dimensional problem we are investigating optical systems to realize the convolution. High speed corollators have been built with bandwidths exceeding 500 MHz, and by coding space into time these corollators may be used to realize the spatial convolution. These are but a few of the possible ways to reduce the current high cost per estimate which results from using a general-purpose digital computer as a synthesis tool. It is clear that the computational problem which is involved in realizing an optimal nonlinear filter is the same as that posed by the numerical solution of partial differential equations, and as such is important for this reason alone.

We are optimistic though, that in five years real systems will be built which depend on numerical solutions to the nonlinear filtering problem, and that perhaps eventually such realizations will become as common place as the current application of extended Kalman-Bucy filters.

BIBLIOGRAPHY

[1] M. ABRAMOWITZ and I.A. STEGUN, *Handbook of Mathematical Functions*, Dover, New York, 1965.

[2] D.L. ALSPACH, « A Bayesian Approximation Technique for Estimation and Control of Time Discrete Stochastic Systems », Ph.D. Dissertation, University of California, San Diego, 1970.

[3] D.L. ALSPACH and H. W. SORENSON, « Approximation of Density Functions by a Sum of Gaussians for Nonlinear Bayesian Estimation », *Proc. Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1970, 19-31.

[4] J.L. BAER and D.P. BOVET, « Compilation of Arithmetic Expressions for Parallel Computations », *Information Processing* 1968, North-Holland Pub. Co., Amsterdam, 1969, 340-346.

[5] R.W. BASS, V.D. NORUM, and L. SCHWARTZ, « Optimal Multichannel Non-Linear Filtering », *J. Math. Anal. Appl.* 16 1966, 152-164.

[6] W.J. BOUKNIGHT, et al., « The Illiac-IV Systems », *Proc. IEEE*, 60 1972, 369-388.

[7] R.S. BUCY, « Linear and Non-linear Filtering », Invited Paper *Proc. IEEE*, 58, 1970, 854-864.

[8] R.S. BUCY, « Realization of Non-Linear Filters », *Proc. Second Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1971, 51-58.

[9] R.S. BUCY, « Building and Evaluating Non-Linear Filters », To appear, *Proc. Symp. on Appl. Math.; Stochastic Diff. Eqns.*, Amer. Math. Soc., April 1972.

[10] R.S. BUCY, C. HECHT and K.D. SENNE, « An Engineer's Guide to Building Nonlinear Filters », *Frank J. Seiler Res. Lab. Rept.*, SRL-TR-72-0004.

[11] R.S. BUCY, C. HECHT and K.G. SENNE, « Optimal Phase Demodulation via Discrete Nonlinear Filtering », Air Force Weapons Laboratory Computer Films No. 72-0401-01, April 1972.

[12] R.S. BUCY and P.D. JOSEPH, *Filtering for Stochastic Processes with Applications to Guidance*, Wiley Interscience, New York, 1968.

[13] R.S. BUCY, M.J. MERRITT and D.S. MILLER, « Hybrid Computer Synthesis of Optimal Discrete Nonlinear Filters », *Proc. Second Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1971, 59-87.

[14] R.S. BUCY and K.D. SENNE, « Realization of Optimum Discrete-Time Nonlinear Estimators », *Proc. Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1970, 6-17.

[15] R.S. BUCY and K.D. SENNE, « Digital Synthesis of Nonlinear Filters, *Automatica* 7 1971, 287-298.

[16] R.S. BUCY and K.D. SENNE, « A Two-Dimensional Passive Ranging Experiment using Optimal Nonlinear Filtering », Air Force Weapons Laboratories Computer Films No. 71-0330-02, March 1971.

[17] J.L. CENTER, « Practical Nonlinear Filtering of Discrete Observations by Generalized Least Squares Approximation of the Conditional Probability Distribution », *Proc. Second. Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1971, 88-99.

[18] Control Data 6400/6500/6600 Computer Systems Reference Manual, CDC Publication 60100000, St. Paul, 1968, Chaps 2 and 3.

[19] H. CRAMER, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1951, pp. 416-451.

[20] W.F. DENAHM and S. PINES, « Sequential Estimation when Measurement Function Nonlinearity is Comparable to Measurement Error », *AIAA J.* 4 1966 1071-1076.

[21] J.L. DOOB, « Heuristic Approach to the Kolmogorov-Smirnov Theorems », *Annals of Mathematical Statistics*, 20 1949, 393-403.

[22] R.J.P. de FIGUEIREDO and Y.G. JAN, « Spline Filters », *Proc. Second Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1971, 88-99.

[23] C. HECHT, « Digital Realization of Non-Linear Filters », *Proc. Second Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1971, 152-158.

[24] C. HECHT, « Synthesis and Realization of Nonlinear Filters », Ph.D. Dissertation, University of Southern California, 1972.

[25] F.B. HILDEBRAND, *Introduction to Numerical Analysis*, McGraw Hill, New York, 1956.

[26] Y.G. JAN, Ph.D. Dissertation, Rice University, 1971.

[27] A.H. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[28] R.M. KELLER, « On Maximally Parallel Schemata », *IEEE Conf. Rec. 11th Annual Symp. on Switching Theory and Automata Theory*, 1970, 32-50.

[29] R.E. LARSON and E. TSE, « Modal Estimation and Parallel Computers », *Proc. Second Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, 1971, 188-197.

[30] J.T. LO, « Finite Dimensional Sensor Orbits and Nonlinear Filtering », Ph.D. Dissertation, University of Southern California, 1969.

[31] A.J. MALLINCKRODT, R.S. BUCY and S.Y. CHENG, « Final Project Report for a Design Study for an Optimal Non-linear Receiver/Demodulator », NASA Contract NAS5-10789, Goddard Space Flight Center, Maryland, 1970.

[32] F.J. MASSEY, Jr., « A note on the Estmiation of a Distribution Function by Confidence Limits », *Annals of Mathematical Statistics*, 21 1950, 116-119.

[33] W.C. MEILANDER, « The Associative Processor in Aircraft Collision Prediction », *NAECON Proc.*, 1968.

[34] D.S. MILLER, « Hybrid Synthesis of Optimal Discrete Nonlinear Filters », Ph.D. Dissertation, University of Southern California, 1971.

[35] K.D. SENNE, « Bayes Law Implementation : Optimal Discrete-Time Phase Estimation », *Proc. SWIEEECO Conf.*, Dallas, April 1972.

[36] K.D. SENNE, « Computer Experiments with Nonlinear Filters », *Eroc. Second Symp. on Nonlinear Estimation and its Applications*, San Diego, 1971, 314-324 (Misprinted — see Appendix of [10]

[37] K.D. SENNE and R.S. BUCY, « Digital Realization of Optimal Discrete-Time Nonlinear Estimators », *Proc. Fourth Annual Princeton Conf. on System Sciences*, Princeton, March 1970, 280-284.

[38] H.W. SORENSON and D.L. ALSPACH, « Recursive Bayesian Estimation using Gaussian Sums », *Automatica* 7 (1971), 465-479.

[39] H.W. SORENSON and A.R. STUBBERUD, « Non-Linear Filtering by Approximation of the A Posteriori Density », *International J. Control*, 8 (1968), 33-51.

[40] K. SRINIVASAN, « State Estimation by Orthogonal Expansion of Probability Distributions, *IEEE Trans. Auto. Control*, AC-15 (1970), 3-10.

[41] E. TSE, « Parallel Computation of the Conditional Mean State Estimate for Nonlinear Systems », *Proc. Second Symp. on Nonlinear Estimation Theory and its Applications*, San Diego, Sept. 1971, 385-394.

[42] A.J. VITERBI, *Principles of Coherent Communication*, McGraw Hill, New York, 1966.

[43] H.L. WEINERT and T. KAILATH, « Recursive Spline Interpolation and Least-Squares Estimation », submitted to Amer. Marth. Soc., 1971.

[44] N. WIENER, *The Fourier Integral and Certain of its Applications*, Cambridge, Cambridge University Press, 1933 (Alson : New York, Dover, 1958).